运筹与管理科学丛书 5

非光滑优化

高岩 著


科学出版社
www.sciencep.com

(O-3106.0101)

ISBN 978-7-03-021502-4

9 787030 215024 >



销售分类建议：高等数学；管理科学

定　价：39.00元

运筹与管理科学丛书　5

非光滑优化

高　岩　著

科学出版社

北　京

内 容 简 介

本书旨在系统介绍非光滑优化理论与方法，全书共分为九章. 第 1 章和第 2 章分别介绍凸集和凸函数的概念和有关性质；第 3 章引入凸函数的次微分，给出凸函数的极值条件与中值定理，并介绍次微分的性质和特殊凸函数的次微分表达式；第 4 章介绍局部 Lipschitz 函数的广义梯度，给出极大值函数广义 Jacobi 的计算；第 5 章阐述拟可微函数及拟微分的定义和性质；第 6 章针对凸规划、Lipschitz 优化、拟可微优化给出最优性条件；第 7 章提出非光滑优化算法，包括下降方法、凸规划的次梯度法、凸规划的割平面法；第 8 章研究非光滑方程组及非线性互补问题；第 9 章介绍非光滑理论在控制论中的应用.

本书可作为应用数学、运筹学与控制论及经济管理有关专业的高年级本科生或研究生教材，也可供相关专业的科研工作者参考.

图书在版编目(CIP)数据

非光滑优化/高岩著. —北京：科学出版社，2008

(运筹与管理科学丛书；5)

ISBN 978-7-03-021502-4

I. 非… II. 高… III. 光滑化(数学) IV. O189

中国版本图书馆 CIP 数据核字 (2008) 第 041433 号

责任编辑：赵彦超／责任校对：陈玉凤

责任印制：赵德静／封面设计：陈　敬

科学出版社出版

北京东黄城根北街 16 号

邮政编码：100717

http://www.sciencep.com

新蕾印刷厂印刷

科学出版社发行　各地新华书店经销

*

2008 年 5 月第 一 版　开本：B5(720 × 1000)

2008 年 5 月第一次印刷　印张：13 1/4

印数：1—3 000　字数：245 000

定价：39.00 元

(如有印装质量问题，我社负责调换〈环伟〉)

《运筹与管理科学丛书》序

运筹学是运用数学方法来刻画、分析以及求解决策问题的科学. 运筹学的例子在我国古已有之, 春秋战国时期著名军事家孙膑为田忌赛马所设计的排序就是一个很好的代表. 运筹学的重要性同样在很早就被人们所认识, 汉高祖刘邦在称赞张良时就说道: "运筹帷幄之中, 决胜千里之外."

运筹学作为一门学科兴起于第二次世界大战期间, 源于对军事行动的研究. 运筹学的英文名字 Operational Research 诞生于 1937 年. 运筹学发展迅速, 目前已有众多的分支, 如线性规划、非线性规划、整数规划、网络规划、图论、组合优化、非光滑优化、锥优化、多目标规划、动态规划、随机规划、决策分析、排队论、对策论、物流、风险管理等.

我国的运筹学研究始于 20 世纪 50 年代, 经过半个世纪的发展, 运筹学队伍已具相当大的规模. 运筹学的理论和方法在国防、经济、金融、工程、管理等许多重要领域有着广泛应用, 运筹学成果的应用也常常能带来巨大的经济和社会效益. 由于在我国经济快速增长的过程中涌现出了大量迫切需要解决的运筹学问题, 因而进一步提高我国运筹学的研究水平、促进运筹学成果的应用和转化、加快运筹学领域优秀青年人才的培养是我们当今面临的十分重要、光荣、同时也是十分艰巨的任务. 我相信,《运筹与管理科学丛书》能在这些方面有所作为.

《运筹与管理科学丛书》可作为运筹学、管理科学、应用数学、系统科学、计算机科学等有关专业的高校师生、科研人员、工程技术人员的参考书, 同时也可作为相关专业的高年级本科生和研究生的教材或教学参考书. 希望该丛书能越办越好, 为我国运筹学和管理科学的发展做出贡献.

袁亚湘

2007 年 9 月

前　言

非光滑优化又称不可微优化, 是最优化理论与方法中的一个重要分支. 所谓非光滑优化, 是指目标函数或约束函数中至少有一个不是连续可微 (光滑) 的非线性规划问题. 由于不具有连续可微性质, 传统的基于微分 (梯度) 概念的优化理论和方法已不再适用于非光滑优化问题. 对经典的微分概念进行推广, 建立各种广义微分概念, 基于广义的微分理论建立相应的最优性理论和算法, 正是非光滑优化研究之所在.

非光滑优化具有广泛的应用背景. 下面给出几个非光滑优化的例子.

例 1　设 $x_1, \cdots, x_m \in \mathbf{R}^n$ 为 m 个实验数据, 要建立一个线性模型, 即求一个超平面 $H = \{ x \in \mathbf{R}^n \,|\, a^{\mathrm{T}}x = b \}$, 其中 $a \in \mathbf{R}^n$, $b \in \mathbf{R}$, 使得 $x_1, \cdots, x_m$ 尽可能接近 H, 这就引出一个不可微优化问题

$$\min_{(a,b)\in\mathbf{R}^{n+1}} \sum_{k=1}^{m} \left| \sum_{i=1}^{n} a_i x_k^i - b \right|, \tag{1}$$

其中, $a \in \mathbf{R}^n$, $b \in \mathbf{R}$ 为变量, a_i, x_k^i 分别为 a 和 x_k 的第 i 个分量. 易见, 问题 (1) 的目标函数带有绝对值, 是非光滑函数. 但是, 人们以往为了回避非光滑问题的困难, 通常考虑下述问题

$$\min_{(a,b)\in\mathbf{R}^{n+1}} \sum_{k=1}^{m} \left(\sum_{i=1}^{n} a_i x_k^i - b \right)^2, \tag{2}$$

即人们通常所熟知的最小二乘问题. 需要指出的是, 问题 (1) 和问题 (2) 的解一般来讲是不一致的. 对大多数情形而言, 问题 (1) 的解较问题 (2) 的解更加符合实际需要.

例 2　考虑非线性互补问题

$$f(x) \geqslant 0, \quad h(x) \geqslant 0, \quad f(x)^{\mathrm{T}} h(x) = 0, \tag{3}$$

其中

$$f(x) = (f_1(x), \cdots, f_n(x))^{\mathrm{T}}, \quad h(x) = (h_1(x), \cdots, h_n(x))^{\mathrm{T}},$$

$f_i(x)$, $h_i(x)$, $i = 1, \cdots, n$ 均为 $\mathbf{R}^n$ 上的连续可微函数, 求解问题 (3) 可等价地转化为求解如下非光滑方程组

$$\min\{ f_i(x), h_i(x) \} = 0, \quad i = 1, \cdots, n. \tag{4}$$

求解方程组 (4) 也等价于求解如下的优化问题:

$$\min_{x\in\mathbf{R}^n}\sum_{i=1}^{n}(\min\{f_i(x),h_i(x)\})^2, \tag{5}$$

显然, (5) 是一个非光滑优化问题.

还有一种非光滑优化来自于优化问题本身, 即求解非线性规划的罚函数方法. 考虑约束优化问题

$$\begin{aligned} &\min\ f(x),\\ &\text{s.t.}\quad g(x)\leqslant 0, \end{aligned} \tag{6}$$

其中, $f(x)$, $g(x)$ 为 $\mathbf{R}^n$ 上的连续可微函数. 利用罚函数法, 求解约束优化问题 (6) 转化为求解下述无约束优化问题

$$\min_{x\in\mathbf{R}^n} f(x)+M\max\{0,g(x)\}, \tag{7}$$

其中 M 为较大的正数. 由于最后一项 $\max\{0,g(x)\}$ 的原因, 问题 (7) 的目标函数是非光滑的.

尽管非光滑优化有广泛的应用, 然而到目前为止, 我们还没有有效的方法处理一般形式的非光滑优化问题, 只能针对一些特殊形式的非光滑优化问题分别进行研究. 在各种类型的非光滑优化中, 凸规划和 Lipschitz 规划是目前影响最大, 也是最被广泛接受的一类非光滑优化问题. 当然, 凸规划是 Lipschitz 规划的特殊形式.

非光滑优化包括许多内容, 目前国外出版的著作都有自身的一套体系. 本书从非光滑优化基本内容入手, 较详细地介绍了凸函数的次微分、局部 Lipschitz 函数的广义梯度、拟可微函数的拟微分及它们的最优性理论, 最后介绍了在控制理论中的应用.

本书的完成得到了国家自然科学基金 (10671126) 和上海市重点学科建设项目 (T0502) 的资助. 作者在写作过程中结合了多年关于非光滑优化的学习和研究工作, 并参阅了国内外相关文献. 由于作者水平有限, 加之时间仓促, 书中一定有许多不妥和错误之处, 敬请各位专家、同行批评指正.

作 者

2008 年 2 月于上海理工大学

目　录

第1章 凸 集

凸集是非光滑分析与优化中最基本的概念之一, 本书以后各章都与本章的内容有关.

1.1 凸集的基本概念

本节引入凸集的概念, 并给出它的一些基本性质.

1.1.1 凸集与凸组合

定义 1.1.1 设 $S \subset \mathbf{R}^n$, 如果对任意 $x_1, x_2 \in S, 0 \leqslant \lambda \leqslant 1$, 有 $\lambda x_1 + (1-\lambda)x_2 \in S$, 则称 S 为凸集.

由定义可以看出, 所谓凸集就是这样的集合, 它的任意两点的连线都在集合中, 可以说凸集具有明显的几何意义.

例 1.1.1 超平面 $H = \{x \in \mathbf{R}^n | p^{\mathrm{T}} x = \alpha\}$ 是凸集, 其中 p 为 n 维向量, α 为实数. 对任意 $x_1, x_2 \in H, 0 \leqslant \lambda \leqslant 1$, 有

$$p^{\mathrm{T}}(\lambda x_1 + (1-\lambda)x_2) = \lambda p^{\mathrm{T}} x_1 + (1-\lambda) p^{\mathrm{T}} x_2 = \alpha,$$

因此 $\lambda x_1 + (1-\lambda)x_2 \in H$, 根据定义, H 是凸集.

例 1.1.2 设 $x_0 \in \mathbf{R}^n$, $\delta > 0$, 容易验证, 以 x_0 为圆心 δ 为半径的球体:

$$B(x, \delta) = \{x \in \mathbf{R}^n | \, ||x - x_0|| \leqslant \delta\}$$

为 $\mathbf{R}^n$ 中的凸集.

另外, 根据凸集的定义很容易验证, $\mathbf{R}^n$ 中空集 $\varnothing$、全空间、所有子空间都是凸集.

定理 1.1.1 设 I 是任意指标集, $S_i \subset \mathbf{R}^n, i \in I$ 是凸集, 则 S_i 的交 $S = \bigcap\limits_{i \in I} S_i$ 是 $\mathbf{R}^n$ 中凸集.

证明 若 S 为空集或单点集, 结论显然成立. 假设 $x_1, x_2 \in S$, 则 $x_1, x_2 \in S_i, i \in I$. 由于 S_i 是凸集, 则对于 $0 \leqslant \lambda \leqslant 1$, 有

$$\lambda x_1 + (1-\lambda)x_2 \in S_i, \quad i \in I,$$

故

$$\lambda x_1 + (1-\lambda)x_2 \in \bigcap_{i\in I} S_i,$$

所以 S 是凸集. 定理得证.

定义 1.1.2 设 $x_1,\cdots,x_m \in \mathbf{R}^n, \lambda_i \geqslant 0,\ i=1,\cdots,m,\ \sum_{i=1}^{m}\lambda_i = 1$, 则点 $x = \sum_{i=1}^{m}\lambda_i x_i$ 称为 $x_1,\cdots,x_m$ 的一个凸组合.

凸组合是凸分析中一个重要概念, 它与凸集有密切联系. 定义 1.1.1 意味着凸集就是 "其中任意两点的凸组合仍属于它本身的集合". 而实际上, 我们也可以通过任意有限点的凸组合来定义凸集, 下面的定理就刻画了这样一个事实.

定理 1.1.2 $S \subset \mathbf{R}^n$ 是凸集的充要条件是 S 中所有元素的凸组合还在 S 中.

证明 设 S 是凸集, $x_1,\cdots,x_m \in S$, 我们将证明 $x_1,\cdots,x_m$ 的凸组合属于 S. 对 m 用数学归纳法. 当 $m=1$ 时, 结论显然成立; 当 $m=2$ 时, 由凸集的定义, 结论也成立. 设结论在 $m \leqslant k$ 时成立, 要证明对于 $\lambda_i \geqslant 0, i=1,\cdots,k+1$ 满足

$$\sum_{i=1}^{k+1}\lambda_i = 1,$$

如果

$$x_i \in S, \quad i=1,\cdots,k+1,$$

则

$$x = \sum_{i=1}^{k+1}\lambda_i x_i \in S.$$

不失一般性, 假设 $\lambda_i > 0, i=1,\cdots,1+k$, 这时

$$1-\lambda_{k+1} = \sum_{i=1}^{k}\lambda_i > 0.$$

根据归纳法假设

$$y = \frac{1}{1-\lambda_{k+1}}x_1 + \cdots + \frac{\lambda_k}{1-\lambda_{k+1}}x_k \in S,$$

这是因为

$$\sum_{i=1}^{k}\frac{\lambda_i}{1-\lambda_{k+1}} = 1,$$

上式为 $x_1,\cdots,x_k$ 的凸组合. 再由凸集的定义, 有

$$x = (1-\lambda_{k+1})y + \lambda_{k+1}x_{k+1} \in S.$$

另一方面, 设集合 S 中元素的所有凸组合都在 S 自身中, 则 S 的任意两个元素的凸组合也在 S 中, 于是 S 是凸集. 定理得证.

定义 1.1.3 $\mathbf{R}^n$ 中集合 S 的凸包是由 S 中的一切凸组合形成的集合, 记为 $\mathrm{co}S$, 换言之, $x \in \mathrm{co}S$ 当且仅当 x 可表示为 $x = \sum\limits_{i=1}^{k} \lambda_i x_i$, 其中 $x_i \in S, \lambda_i \geqslant 0, i = 1, \cdots, k, \sum\limits_{i=1}^{k} \lambda_i = 1$, k 为一正整数.

很容易验证, S 的凸包是包含 S 的最小凸集. 事实上, 不难验证它是包含 S 的所有凸集的交集. 凸包也是对一个给定非凸集合进行凸化的手段.

$\mathbf{R}^n$ 中有限点集 $\{\alpha_1, \cdots, \alpha_m\}$, 其中 $\alpha_i \in \mathbf{R}^n, i = 1, \cdots, m$ 的凸包由形如 $\lambda_1\alpha_1 + \cdots + \lambda_m\alpha_m$ 的向量构成, 其中 $\lambda_i \geqslant 0$, $i = 1, \cdots, m$, 且有 $\sum\limits_{i=1}^{m} \lambda_i = 1$, 亦可表示为

$$\left\{ \sum_{j=1}^{m} \lambda_i a_i | \lambda_i \geqslant 0,\ i = 1, \cdots, m \right\},$$

它是 $\mathbf{R}^n$ 空间中的一个凸多面体.

由定义 1.1.3 知, 凸包 $\mathrm{co}S$ 是由 S 的所有有限多个点的凸组合构成的集合, 但定义 1.1.3 没有对构成这个凸组合所需的点数给出任何限制, 实际上, 对于 n 维空间中的集合 S, 只需至多 $n+1$ 个点的凸组合就可以表示 $\mathrm{co}S$ 中的点, 下面的 Caraheodory 定理就揭示了这样的事实.

定理 1.1.3 (Caraheodory 定理) 设 $S \subset \mathbf{R}^n$, 则 S 的凸包 $\mathrm{co}S$ 中的任意一点可以表示成 S 中至多 $n+1$ 个点的凸组合, 即对任意 $x \in \mathrm{co}S$, 存在常数 $r \leqslant n+1$ 以及 $x_i \in S, \lambda_i \geqslant 0, i = 1, \cdots, r$ 满足 $\sum\limits_{i=1}^{r} \lambda_i = 1$, 使得

$$x = \sum_{i=1}^{r} \lambda_i x_i. \tag{1.1.1}$$

证明 根据定理 1.1.2, 只要证明 $r \leqslant n+1$ 即可. 以下证明, 如果 $r > n+1$, 则式 (1.1.1) 右边的非零项可以减少. 不妨假设

$$\lambda_i > 0, \quad i = 1, \cdots, r, \quad r > n+1.$$

取 $n+1$ 维向量 $(x_i, 1)$, $i = 1, \cdots, r$, 因为向量的个数 $r > n+1$, 因此线性相关, 故存在不全为零的常数 α_i, $i = 1, \cdots, r$, 使得

$$\sum_{i=1}^{r} \alpha_i (x_i, 1) = 0,$$

于是有

$$\sum_{i=1}^{r}\alpha_i x_i = 0, \quad \sum_{i=1}^{r}\alpha_i = 0. \tag{1.1.2}$$

由

$$\sum_{i=1}^{r}\alpha_i = 0$$

可知, α_i, $i=1,\cdots,r$ 中一定存在正数, 记

$$\varepsilon_0 = \min\left\{\frac{\lambda_i}{\alpha_i}|\alpha_i > 0,\ i=1,\cdots,r\right\},$$

于是存在一个 i_0, 使得

$$\varepsilon_0 = \frac{\lambda_{i_0}}{\alpha_{i_0}},$$

进而有

$$\bar{\lambda}_i = \lambda_i - \varepsilon_0\alpha_i \geqslant 0, \quad i=1,\cdots,r, \tag{1.1.3}$$

特别地, $\bar{\lambda}_{i_0}=0$. 由式 (1.1.2) 得

$$\sum_{i=1}^{r}\bar{\lambda}_i x_i = \sum_{i=1}^{r}\lambda_i x_i - \varepsilon_0\sum_{i=1}^{r}\alpha_i x_i = x,$$

$$\sum_{i=1}^{r}\bar{\lambda}_i = \sum_{i=1}^{r}\lambda_i - \varepsilon_0\sum_{i=1}^{r}\alpha_i = 1, \tag{1.1.4}$$

这说明, x 还可以表示为式 (1.1.1) 的形式, 但却减少了一项 (在式 (1.1.4) 中 $\bar{\lambda}_{i_0}=0$). 定理得证.

1.1.2　代数运算

在非光滑分析中, 通常的集合加法和数乘运算按如下法则, 通常也称为 Minkowski 加法和数乘.

定义 1.1.4　设 $S_1, S_2 \subset \mathbf{R}^n$, $\lambda \in \mathbf{R}^n$, 则 $\lambda S_1 = \{\lambda x | x \in S_1\}$ 称为集合 S_1 和 λ 的数乘;

$$S_1 + S_2 = \{x_1 + x_2 | x_1 \in S_1,\ x_2 \in S_2\}$$

称为 S_1 和 S_2 的和.

下述结论显然成立.

命题 1.1.1　设 $S_1, S_2 \subset \mathbf{R}^n$ 为凸集, $\lambda \in \mathbf{R}$, 则 λS_1 和 $S_1 + S_2$ 也为凸集.

定理 1.1.4　设 S 为 $\mathbf{R}^n$ 中凸集, $\lambda_1 \geqslant 0, \lambda_2 \geqslant 0$, 则有

$$(\lambda_1 + \lambda_2)S = \lambda_1 S + \lambda_2 S. \tag{1.1.5}$$

证明 当 $\lambda_1 = \lambda_2 = 0$ 时, 式 (1.1.5) 显然成立. 设 $\lambda_1 + \lambda_2 > 0$, 容易验证, 即使集合 S 不是凸, 下述性质也成立:

$$(\lambda_1 + \lambda_2)S \subset \lambda_1 S + \lambda_2 S. \tag{1.1.6}$$

对于凸集 S, 容易验证

$$\lambda S + (1-\lambda)S \subset S,$$

其中 $0 \leqslant \lambda \leqslant 1$, 于是有

$$\frac{\lambda_1}{\lambda_1+\lambda_2}S + \frac{\lambda_2}{\lambda_1+\lambda_2}S \subset S,$$

上式两边乘以 $\lambda_1 + \lambda_2$ 得

$$\lambda_1 S + \lambda_2 S \subset (\lambda_1 + \lambda_2)S, \tag{1.1.7}$$

联立式 (1.1.6) 和式 (1.1.7) 得式 (1.1.5). 定理得证.

1.1.3 凸锥

在凸集中, 一个比较重要的特殊情形是凸锥. 凸锥是非光滑分析和优化中研究的重要对象之一.

定义 1.1.5 设 S 为 $\mathbf{R}^n$ 中集合, $x_0 \in S$, 且对任意 $\lambda > 0$, 有 $x_0 + \lambda(x - x_0) \in S$, 则称 S 是以 x_0 为顶点的锥. 特别当 S 为凸集时, S 称为凸锥.

在最优化研究中, 人们最关心的是以 0 为顶点的锥, 以后除特殊声明, 所提到的锥均指以 0 为顶点的锥. 易见, $S \subset \mathbf{R}^n$ 是锥的充要条件是对任意 $\lambda > 0$, 有 $\lambda S = S$.

例 1.1.3 下述两个集合是 $\mathbf{R}^n$ 中的凸锥:

$$S_1 = \{(x_1, \cdots, x_n)^{\mathrm{T}} | x_i \geqslant 0,\ i = 1, \cdots, n\},$$

$$S_2 = \{(x_1, \cdots, x_n)^{\mathrm{T}} | x_i > 0,\ i = 1, \cdots, n\}.$$

定理 1.1.5 $S \subset \mathbf{R}^n$ 是凸锥的充分必要条件是它对加法和正数乘法封闭.

证明 设 S 是凸锥, 由于是锥, 它对正数乘法是封闭的, 又 S 是凸集, 则对任意 $x_1, x_2 \in S$, 有

$$z = \frac{1}{2}x_1 + \frac{1}{2}x_1 \in S,$$

所以 $x_1 + x_2 = 2z \in S$, 即 S 关于加法封闭.

设 S 对加法和正数乘法封闭. 由于对正数乘法封闭, 所以 S 是锥, 故对任意

$$x_1, x_2 \in S, \quad 0 < \lambda < 1,$$

有

$$\lambda x_1 \in S, \quad (1-\lambda)x_2 \in S.$$

而由 S 对加法封闭性, 得 $\lambda x_1 + (1-\lambda)x_2 \in S$, 所以 S 是凸集, 进而是凸锥. 定理得证.

由定理 1.1.5 易得下面的推论.

推论 1.1.1　$S \subset \mathbf{R}^n$ 是凸锥的充分必要条件是 S 包含它的元素的全部正线性组合, 即对任意 $x_i \in S$, $\lambda > 0$, $i = 1, \cdots, m$, 有

$$\lambda_1 x_1 + \cdots + \lambda_m x_m \in S.$$

定义 1.1.6　设 $S \subset \mathbf{R}^n$ 是凸锥, 它的极锥定义如下:

$$S^\circ = \{d \in \mathbf{R}^n | d^{\mathrm{T}} x \leqslant 0,\ \forall x \in S\}.$$

从极锥的定义不难看出, S° 是闭凸锥. 不难验证: 如果 S 是一个子空间, S° 是它的正交补; 如果 $S_1 \subset S_2$, 则 $S_1^\circ \supset S_2^\circ$; 锥与它的极锥交非空, 即 $S \bigcap S^\circ = \{0\}$.

例 1.1.4　设集合 S 为有限点形成的凸锥, 即

$$S = \left\{ \sum_{j=1}^{m} \lambda_i a_i | \lambda_i \geqslant 0,\ i = 1, \cdots, m \right\},$$

其中 $a_i \in \mathbf{R}^n, i = 1, \cdots, m$, 则 S 的极锥为

$$S^\circ = \{y \in \mathbf{R}^n | y^{\mathrm{T}} x_i \leqslant 0, i = 1, \cdots, m\}.$$

1.2　凸集上的投影

点到子空间的投影算子有很多好的性质, 例如线性、对称性、半正定性、幂等性、非膨胀性等. 本节讨论点到凸集的投影, 它在平衡问题以及变分不等式问题中有许多重要应用.

给定非空闭集 $S \subset \mathbf{R}^n$ 和固定的 $x \in \mathbf{R}^n$, 考虑下述问题:

$$\inf_{y \in S} \frac{1}{2} ||y - x||^2, \tag{1.2.1}$$

它是寻找 x 到集合 S 中的最近点, 即投影. 固定 $x \in \mathbf{R}^n$, 定义下述函数:

$$f_x(y) = \frac{1}{2} ||y - x||^2, \quad y \in \mathbf{R}^n, \tag{1.2.2}$$

给定 $s \in S$, 考虑水平集

$$L_s = \{y \in \mathbf{R} | f_x(y) \leqslant f_x(s)\}.$$

显然, 问题 (1.2.1) 与问题 $\inf\limits_{y \in S \cap L_s} f_x(y)$ 等价. 由于函数 $f_x(y)$ 连续且非负, 则 $S \bigcap L_s$ 为紧集, 于是点 x 到集合 S 的最近点, 即投影是存在的. 这样式 (1.1.2) 中的下确界 inf 可用极小值 min 来代替.

事实上, 投影的存在性并不需要集合的凸性来保证, 但是凸性却可以保证它的唯一性. 假设 y_1 和 y_2 都是问题 (1.2.1) 的解, 令 $x_i = y_i - x$, $i = 1, 2$, 记 $y_0 = \frac{1}{2}(y_1 + y_2)$, 利用关系式

$$\frac{1}{2}||y_1 + y_2||^2 = ||y_1||^2 + ||y_2||^2 - \frac{1}{2}||y_2 - y_1||^2,$$

得

$$f_x(y_0) = \frac{1}{2}(f_x(y_1) + f_x(y_2)) - \frac{1}{8}||y_2 - y_1||^2.$$

由 y_1 和 y_2 均为 (1.2.1) 的解和 $y_0 \in S$, 则有 $y_1 = y_2$, 这就说明了凸集投影的唯一性.

记 $p_S(x)$ 为点 x 到集合 S 上的投影, 我们有下述定理.

定理 1.2.1 设 $S \subset \mathbf{R}^n$ 为凸集, $x \in \mathbf{R}^n$, 则 $y_x \in S$ 为 x 到 S 的投影 $p_S(x)$ 的充要条件是

$$(x - y_x)^{\mathrm{T}}(y - y_x) \leqslant 0, \quad \forall\, y \in S. \tag{1.2.3}$$

证明 必要性. 设 S 为凸集. 注意到 y_x 是问题 (1.2.1) 的解, 取任意的 $y \in S$, 根据 S 的凸性有

$$y_x + \lambda(y - y_x) \in S, \quad 0 < \lambda < 1,$$

由式 (1.2.2) 可得

$$\begin{aligned}\frac{1}{2}||y_x - x||^2 &= f_x(y_x)\\ &\leqslant f_x(y_x + \alpha(y - y_x))\\ &= \frac{1}{2}||y_x - x + \lambda(y - y_x)||^2.\end{aligned}$$

展开平方项并整理得

$$0 \leqslant \lambda(y_x - x)^{\mathrm{T}}(y - y_x) + \frac{1}{2}\lambda^2||y - y_x||^2,$$

两边除以 λ, 并令 $\lambda \to 0^+$, 即得式 (1.2.3).

充分性. 假设 $y_x \in S$ 满足式 (1.2.3), 如果 $y_x = x$, 则 y_x 显然是 (1.2.1) 的解. 考虑 $y_x \neq x$, 对任意的 $y \in S$, 利用 Cauchy-Schwarz 不等式, 直接推导得

$$\begin{aligned}0 &\geqslant (x-y_x)^{\mathrm{T}}(y-y_x)\\&=(x-y_x)^{\mathrm{T}}(y-x+x-y_x)\\&=\|x-y_x\|^2+(x-y_x)^{\mathrm{T}}(y-x)\\&\geqslant \|x-y_x\|^2-\|x-y\|\,\|x-y_x\|.\end{aligned}$$

注意到 $\|x-y_x\|>0$, 两边除以 $\|x-y_x\|$ 知 y_x 是式 (1.2.3) 的解. 定理得证.

进一步, 我们有下述定理.

定理 1.2.2　设 $S\subset\mathbf{R}^n$ 为凸集, 对任意 $x_1,x_2\in\mathbf{R}^n$, 下式成立:

$$\|p_S(x_1)-p_S(x_2)\|^2\leqslant(p_S(x_1)-p_S(x_2))^{\mathrm{T}}(x_1-x_2). \tag{1.2.4}$$

证明　在式 (1.2.3) 中, 取

$$x=x_1,\quad y=p_S(x_2)\in S,$$

则有

$$(p_S(x_2)-p_S(x_1))^{\mathrm{T}}(x_1-p_S(x_1))\leqslant 0.$$

类似地有

$$(p_S(x_1)-p_S(x_2))^{\mathrm{T}}(x_2-p_S(x_2))\leqslant 0.$$

上述两式相加得

$$(p_S(x_1)-p_S(x_2))^{\mathrm{T}}(x_2-x_1+p_S(x_1)-p_S(x_2))\leqslant 0,$$

故式 (1.2.3) 成立. 定理得证.

由定理 1.2.2, 我们立刻得到下述两个有趣的结论:

$$0\leqslant(p_S(x_1)-p_S(x_2))^{\mathrm{T}}(x_1-x_2),\quad \forall x_1,x_2\in\mathbf{R}^n,$$

即投影算子 p_S 的单调性. 结合 Cauchy-Schwarz 不等式得到

$$\|p_S(x_1)-p_S(x_2)\|\leqslant\|x_1-x_2\|, \tag{1.2.5}$$

即投影算子 p_S 的非膨胀性. 特别是如果 $0\in S$, 则有 $\|p_S(x)\|\leqslant\|x\|$.

定理 1.2.3　设 $S\subset\mathbf{R}^n$ 是闭凸锥, 则 y_x 是点 x 到 S 投影的充要条件是

$$y_x\in S,\quad x-y_x\in S^{\circ},\quad (x-y_x)^{\mathrm{T}}y_x=0. \tag{1.2.6}$$

证明　必要性. 设 y_x 是 x 到 S 的投影, 根据定理 1.2.1, 有

$$(x-y_x)^{\mathrm{T}}(y-y_x)\leqslant 0,\quad \forall y\in S. \tag{1.2.7}$$

将 $y=\alpha y_x$, $\alpha\geqslant 0$, 代入式 (1.2.7) 得

$$(\alpha-1)(x-y_x)^{\mathrm{T}}y_x\leqslant 0,$$

由于 $\alpha-1$ 可取正值或负值, 于是有

$$(x-y_x)^{\mathrm{T}}y_x=0,$$

这样式 (1.2.7) 变为

$$y^{\mathrm{T}}(x-y_x)\leqslant 0,\quad \forall y\in S,$$

即

$$x-y_x\in S^{\circ}.$$

充分性. 设 y_x 满足式 (1.2.6). 对任意 $y\in S$, 由式 (1.2.2) 的记号, 有

$$\begin{aligned}f_x(y)&=\frac{1}{2}||x-y_x+y_x-y||^2\geqslant f_x(y)\\&=f_x(y_x)+(x-y_x)^{\mathrm{T}}(y_x-y).\end{aligned}$$

又由式 (1.2.6) 可得

$$(x-y_x)^{\mathrm{T}}(y_x-y)=-(x-y_x)^{\mathrm{T}}\geqslant 0,$$

于是

$$f_x(y)\geqslant f_x(y_x),$$

故 y_x 是式 (1.2.1) 的解. 定理得证.

1.3　凸集的分离定理

平面中两个不相交凸集的一个明显几何性质, 就是存在一条直线将它们分开, 使得一个集合在直线一侧, 另外一个集合在直线的另一侧. 这一事实在一般的 n 维空间中就是所谓的凸集分离定理. 凸集的分离定理在最优化理论, 特别是在最优性条件的建立中起着重要的作用. 本节介绍几个凸集分离定理.

1.3.1　分离定理

首先介绍单点集与凸集的分离定理.

定理 1.3.1 (分离定理)　设 $S\subset\mathbf{R}^n$ 是非空闭凸集, $x\in\mathbf{R}^n$ 且 $x\notin S$, 则存在 $p\in\mathbf{R}^n$, 使得

$$\sup_{y\in S}p^{\mathrm{T}}y<p^{\mathrm{T}}x. \tag{1.3.1}$$

证明　令

$$p = x - p_S(x) \neq 0,$$

其中 $p_S(x)$ 为点 x 到集合 S 的投影, 根据投影性质 (见定理 1.2.1) 得

$$\begin{aligned} 0 &\geqslant (x - p_S(x))(y - p_S(x)) \\ &= p^{\mathrm{T}}(y - x + p) \\ &= p^{\mathrm{T}}y - p^{\mathrm{T}}x + \|p\|^2, \end{aligned}$$

于是

$$p^{\mathrm{T}}x - \|p\|^2 \geqslant p^{\mathrm{T}}y, \quad \forall y \in \mathrm{S}.$$

注意到 $\|p\| \geqslant 0$, 故

$$p^{\mathrm{T}}x \geqslant p^{\mathrm{T}}y, \quad \forall y \in S,$$

这样就得到式 (1.3.1). 定理得证.

如果在式 (1.3.1) 中, 以 $-p$ 代替 p, 则定理 1.3.1 可叙述为: 存在 $p \in \mathbf{R}^n$, 使得 $p^{\mathrm{T}}x < \inf\limits_{y\in S} p^{\mathrm{T}}y$. 显然在式 (1.3.1) 中, $p \neq 0$, 因此我们可以要求 p 为单位向量, 即 $\|p\| = 1$. 定理 1.3.1 说明存在一个超平面, 将一点和一个凸集分为两部分.

定理 1.3.2 (分离定理)　设 S_1, S_2 为 $\mathbf{R}^n$ 中两个非空闭凸集, 且 $S_1 \bigcap S_2 \neq \varnothing$, 如果 S_2 是有界的, 则存在 $p \in \mathbf{R}^n$, 使得

$$\sup_{y\in S_1} p^{\mathrm{T}}y < \min_{y\in S_2} p^{\mathrm{T}}y. \tag{1.3.2}$$

证明　集合 $S_1 - S_2$ 是凸的, 又 S_2 是紧的, 因此也是闭的. $S_1 \bigcap S_2 = \varnothing$ 意味着 $0 \notin S_1 - S_2$, 于是将定理 1.3.1 应用到点 0 和集合 $S_1 - S_2$ 上, 则存在 $p \in \mathbf{R}^n$, 使得

$$\sup_{y\in S_1 - S_2} p^{\mathrm{T}}y < p^{\mathrm{T}}0 = 0. \tag{1.3.3}$$

注意到

$$\begin{aligned} \sup_{y\in S_1 - S_2} p^{\mathrm{T}}y &= \sup_{y\in S_1} p^{\mathrm{T}}y + \sup_{y\in S_2} p^{\mathrm{T}}(-y) \\ &= \sup_{y\in S_1} p^{\mathrm{T}}y - \inf_{y\in S_2} p^{\mathrm{T}}y. \end{aligned} \tag{1.3.4}$$

结合式 (1.3.3) 和式 (1.3.4), 再注意到 S_2 是紧集, 式 (1.3.4) 中最后一项中 inf 可用 min 代替, 得到式 (1.3.2). 定理得证.

当然式 (1.3.2) 也可等价地表示为

$$\max_{y\in S_2} p^{\mathrm{T}}y < \inf_{y\in S_1} p^{\mathrm{T}}y.$$

前面讨论的分离定理要求其中的一个集合是有界的, 下面我们去掉有界性假设, 讨论一般形式的分离定理.

定理 1.3.3 (分离定理) 设 $S_1, S_2 \subset \mathbf{R}^n$ 为非空凸集且 $S_1 \bigcap S_2 \neq \varnothing$, 则存在 $p \in \mathbf{R}^n$, 使得

$$\sup_{y\in S_1} p^{\mathrm{T}} y < \inf_{y\in S_2} p^{\mathrm{T}} y. \tag{1.3.5}$$

证明 令 $S = S_1 - S_2$, 因为 $S_1 \bigcap S_2 \neq \varnothing$, 则 $0 \notin S$. 在定理 1.3.1 中, 取 $x = 0$ 得: 存在 $p \in \mathbf{R}^n$, 使得

$$p^{\mathrm{T}} y < 0, \quad y \in S,$$

即得

$$p^{\mathrm{T}} x_2 \leqslant p^{\mathrm{T}} x_1, \quad \forall x_1 \in S_1, x_2 \in S_2,$$

由 x_1, x_2 的任意性得式 (1.3.5). 定理得证.

1.3.2 Farkas 引理和 Gordan 定理

利用分离定理, 我们可得到两个重要定理: Farkas 定理 (也称 Farkas 引理) 和 Gordan 定理, 它们在光滑和非光滑优化最优性条件建立中起着直接作用.

定理 1.3.4 (Farkas 引理) 设 A 是 $m \times n$ 阶矩阵, $c \in \mathbf{R}^n$, 则下面两组线性不等式恰好一组有解:

$$Ax \leqslant 0, \quad c^{\mathrm{T}} x > 0, \quad x \in \mathbf{R}^n, \tag{1.3.6}$$

$$A^{\mathrm{T}} y = c, \quad y \geqslant 0, \quad y \in \mathbf{R}^m, \tag{1.3.7}$$

其中 $y \geqslant 0$ 代表 y 的每个分量都大于等于 0.

证明 假设不等式 (1.3.7) 有解, 即存在一个 $y \geqslant 0$, 使得 $A^{\mathrm{T}} y = c$. 设 $x \in \mathbf{R}^n$ 满足 $Ax \leqslant 0$, 则

$$c^{\mathrm{T}} x = y^{\mathrm{T}} Ax \leqslant 0,$$

因此不等式组 (1.3.6) 无解.

现在假设 (1.3.7) 无解, 构造集合

$$S = \{x \in \mathbf{R}^n | x = A^{\mathrm{T}} y, y \geqslant 0\}.$$

注意到, S 是闭集且 $c \notin S$, 由定理 1.3.1 可得, 存在一个向量 $p \in \mathbf{R}^n$ 和实数 α, 使得

$$p^{\mathrm{T}} c > \alpha, \quad p^{\mathrm{T}} x \leqslant \alpha, \quad \forall x \in S.$$

因为 $0 \in S$, 则 $\alpha \geqslant 0$, 所以 $p^{\mathrm{T}} c > 0$. 又

$$\alpha \geqslant p^{\mathrm{T}} A^{\mathrm{T}} y = y^{\mathrm{T}} Ap$$

对一切 $y \geqslant 0$ 成立, 令 y 任意大, 故 $\alpha \geqslant y^{\mathrm{T}}Ap$ 蕴涵着 $Ap \leqslant 0$, 向量 p 满足

$$Ap \leqslant 0, \quad p^{\mathrm{T}} > 0, \quad Ap \leqslant 0, \quad p^{\mathrm{T}}c > 0,$$

它是不等式组 (1.3.6) 的解. 定理得证.

Farkas 引理还有其他的一些形式, 在定理 1.3.4 中, 令 $A^{\mathrm{T}} = [A^{\mathrm{T}}, -I]$, 则得到结论: 下面两个线性不等式组恰好一个有解:

$$Ax \leqslant 0, \quad x \geqslant 0, \quad c^{\mathrm{T}}x > 0, \quad x \in \mathbf{R}^n, \tag{1.3.8}$$

$$A^{\mathrm{T}}y \geqslant c, \quad y \geqslant 0, \quad y \in \mathbf{R}^m. \tag{1.3.9}$$

在定理 1.3.4 中, 令

$$A^{\mathrm{T}} = [A^{\mathrm{T}}, B^{\mathrm{T}}, -B^{\mathrm{T}}],$$

其中 B 为 $l \times n$ 阶矩阵, 则得到结论: 下述两个线性不等式组恰好一个有解:

$$Ax \leqslant 0, \quad Bx = 0, \quad c^{\mathrm{T}}x > 0, \quad x \in \mathbf{R}^n, \tag{1.3.10}$$

$$A^{\mathrm{T}}y + B^{\mathrm{T}}z = c, \quad y \geqslant 0, \quad y \in \mathbf{R}^m. \tag{1.3.11}$$

利用分离定理还可以得到下述 Gordan 定理.

定理 1.3.5 (Gordan 定理)　设 A 是一个 $m \times n$ 阶矩阵, 则线性不等式组

$$Ax < 0, \quad x \in \mathbf{R}^n \tag{1.3.12}$$

有解的充分必要条件是线性不等式组

$$A^{\mathrm{T}}y = 0, \quad y \geqslant 0 \tag{1.3.13}$$

无非零解.

证明　首先证明, 如果不等式组 (1.3.12) 有解 $\widehat{x}$, 则不等式组 (1.3.13) 不能有非零解. 用反证法, 假设 $\widehat{y}$ 是 (1.3.13) 的非零解, 则由

$$A\widehat{x} < 0, \quad \widehat{y} \geqslant 0, \quad \widehat{y} \neq 0$$

可得

$$\widehat{y}A\widehat{x} < 0,$$

即

$$\widehat{x}^{\mathrm{T}}A^{\mathrm{T}}\widehat{y} < 0.$$

由于 $A^{\mathrm{T}}\widehat{y} = 0$, 因此

$$\widehat{y}A\widehat{x} = 0,$$

与 $\hat{y}A\hat{x} < 0$ 矛盾, 所以不等式组 (1.3.13) 无非零解.

现在假设不等式组 (1.3.12) 没有解. 考虑下述两个集合:

$$S_1 = \{z \in \mathbf{R}^m | z = Ax, x \in \mathbf{R}^n\},$$

$$S_2 = \{z \in \mathbf{R}^m | z < 0\},$$

由定义, S_1 和 S_2 均为 $\mathbf{R}^m$ 中非空凸集且有 $S_1 \bigcap S_2 \neq \varnothing$. 根据定理 1.3.3, 存在 $\mathbf{R}^m$ 中的非零向量 $p \in \mathbf{R}^m$, 使得

$$p^{\mathrm{T}} Ax \geqslant p^{\mathrm{T}} z, \quad \forall x \in \mathbf{R}^n, \quad z \in S_2.$$

注意到 z 的每个分量都可以取任意大的负数, 因此必有

$$p \geqslant 0.$$

令 $z \to 0$, 对每个 $x \in \mathbf{R}^n$ 必有

$$p^{\mathrm{T}} Ax \geqslant 0,$$

选取 $x = -A^{\mathrm{T}} p$, 得

$$-||A^{\mathrm{T}} p||^2 \geqslant 0,$$

于是有

$$A^{\mathrm{T}} p = 0,$$

因此不等式组 (1.3.11) 有解. 定理得证.

1.4 多面体的极点和极方向

多面体集是一类重要的凸集, 是通过有限个闭半空间的交而形成的凸集, 在各种凸集中多面体集是最简单也最常见的. 本节将讨论多面体, 特别是有关它的极点和极方向以及用极点和极方向表示多面体.

1.4.1 极点和极方向性质

定义 1.4.1 集合

$$S = \{x \in \mathbf{R}^n | p_i^{\mathrm{T}} x \leqslant a_i,\ i = 1, \cdots, m\},$$
$$p_i \in \mathbf{R}^n, \quad \alpha_i \in \mathbf{R}, \quad i = 1, \cdots, m$$

称为 $\mathbf{R}^n$ 中的多面体集.

例如, 下述集合就是 $\mathbf{R}^2$ 中的一个多面体集:

$$S = \{(x_1, x_2\} \in \mathbf{R}^2 | -x_1 + x_2, x_2 \leqslant 4, x_1 \geqslant 0, x_2 \geqslant 0\}.$$

多面体集就是由线性不等式组表示的集合, 通常也表示为下述形式:

$$S=\{x\in\mathbf{R}^n|Ax\leqslant b\}, \tag{1.4.1}$$

或

$$S=\{x\in\mathbf{R}^n|Ax=b,x\geqslant 0\}, \tag{1.4.2}$$

其中 A 为 $m\times n$ 阶矩阵, b 为 m 维向量.

在多面体集研究中有两个重要的概念就是极点和极方向.

定义 1.4.2 $S\subset\mathbf{R}^n$ 是一非空凸集, 如果 $x_1,x_2\in S$, $0\leqslant\lambda\leqslant 1$, $x=\lambda x_1+(1-\lambda)x_2$, 必有 $x=x_1=x_2$, 则称 x 为 S 的极点.

对有界凸多面集, 集合中的任何点都可以表示为极点的凸组合, 然而对于无解情形, 则有所不同, 为了研究无界多面体集, 需要引进极方向的概念.

定义 1.4.3 设 $S\subset\mathbf{R}^n$ 是闭凸集, $d\in\mathbf{R}^n$, 如果对任意 $x\in S$, $\lambda\geqslant 0$, 都有 $x+\lambda d\in S$, 则 d 称为 S 的一个方向. 设 d_1, d_2 为 S 中两个方向, 如果对任意 $\alpha>0$, 有 $d_1\neq\alpha d_2$, 则称 d_1 和 d_2 是不同的方向. 如果 S 中的方向 d 不能表示为 S 中两个不同方向的正线性组合, 即如果 $d=\lambda_1 d_1+\lambda_2 d_2$, λ_1, $\lambda_2>0$, 则对某个 $\alpha>0$, 有 $d_1=\alpha d_2$, 称 d 为 S 的极方向.

例 1.4.1 设

$$S=\{\ (x_1,\ x_2)\in\mathbf{R}^2\,|\,|x_1|\leqslant x_2\,\},$$

S 的方向是所有与向量 $(0,\ 1)^{\mathrm{T}}$ 夹角小于等于 $45°$ 的向量, 特别 $d_1=(1,1)^{\mathrm{T}}$ 和 $d_2=(-1,1)^{\mathrm{T}}$ 是 S 的两个极方向. S 的任何其他方向都可以表示为 d_1 和 d_2 的正线性组合.

考虑由式 (1.4.2) 给出的多面体集, 不妨假设 A 的秩为 m (如果 A 的秩少于 m, 在 $Ax=b$ 中删掉多余的方程), 重新排列 A 的列, 使 $A=[B,N]$, 其中 B 是 $m\times m$ 阶满秩矩阵, N 是 $m\times(n-m)$ 阶矩阵. 设 x_B 和 x_N 为分别对应于 B 和 N 的向量, 此时 $Ax=b$, $x\geqslant 0$ 可表示为

$$Bx_B+Nx_N=b,\quad x_B\geqslant b,\quad x_N\geqslant 0. \tag{1.4.3}$$

下面定理给出了多面体集 S 极点的充要条件.

定理 1.4.1 $x\in\mathbf{R}^n$ 为集合 (1.4.2) 极点的充要条件是 A 可分解为 $A=[B,N]$ 形式, 其中 B 是 $m\times n$ 阶可逆阵, N 是 $m\times(n-m)$ 阶矩阵, 使得

$$x=\begin{pmatrix}x_B\\x_N\end{pmatrix}=\begin{pmatrix}B^{-1}b\\0\end{pmatrix},\quad B^{-1}b\geqslant 0. \tag{1.4.4}$$

证明 充分性. 假设 A 能够分解成 $A=[B,N]$ 形式, 且满足式 (1.4.4), 显然 $x\in S$, 以下证明 x 为 S 的极点. 假设 $x=\lambda_1x_1+\lambda_2x_2$, 其中 $x_1,x_2\in S$, $0\leqslant\lambda\leqslant 1$, 记

$$x_1^{\mathrm{T}}=(x_{11}^{\mathrm{T}},x_{12}^{\mathrm{T}})^{\mathrm{T}},\quad x_2^{\mathrm{T}}=(x_{21}^{\mathrm{T}},x_{22}^{\mathrm{T}})^{\mathrm{T}},$$

则有

$$\begin{pmatrix} B^{-1}b \\ 0 \end{pmatrix}=\lambda\begin{pmatrix} x_{11} \\ x_{12} \end{pmatrix}+(1-\lambda)\begin{pmatrix} x_{21} \\ x_{22} \end{pmatrix}.$$

由于 x_{12}, $x_{22}\geqslant 0$ 和 $0\leqslant\lambda\leqslant 1$, 则有 $x_{12}=x_{22}=0$, 再由式 (1.4.3), 可见 $x_{11}=x_{21}=B^{-1}b$, 即 $x=x_1=x_2$, 故 x 是 S 的一个极点.

必要性. 假设 $x\in S$ 是 S 的一个极点, 不失一般性, 假设 $x=(x_1,\cdots,x_k,0,\cdots,0)^{\mathrm{T}}$, 其中 $x_i>0$, $i=1,\cdots,k$. 记 a_i 为 A 的第 i 列, 我们首先证明 $a_1,\cdots,a_k$ 线性无关. 用反证法, 假定 $a_1,\cdots,a_k$ 线性相关, 则必存在一组不全为零的常数 $\lambda_1,\cdots,\lambda_k$, 使得 $\sum\limits_{i=1}^{k}\lambda_ia_i=0$. 记 $\lambda=(\lambda_1,\cdots,\lambda_k,0,\cdots,0)^{\mathrm{T}}$, 构造下述两个向量:

$$x^{(1)}=x+\alpha\lambda,\quad x^{(2)}=x-\alpha\lambda\frac{-b\pm\sqrt{b^2-4ac}}{2a},$$

其中 $\alpha>0$ 的选择要使 $x^{(1)},x^{(2)}\geqslant 0$. 直接计算得

$$\begin{aligned} Ax^{(1)}&=\sum_{i=1}^{k}(x_i+\alpha\lambda_i)\,a_i \\ &=\sum_{i=1}^{k}x_ia_i+\alpha\sum_{i=1}^{k}\lambda_ia_i \\ &=b. \end{aligned}$$

类似可得

$$Ax^{(2)}=b,$$

因此

$$x^{(1)},\ x^{(2)}\in S.$$

又因为 $\alpha>0$, 所以

$$x^{(1)}\neq x^{(2)},$$

而

$$x=\frac{1}{2}x^{(1)}+\frac{1}{2}x^{(2)},$$

这与 x 是极点相矛盾, 于是 $a_1, \cdots, a_k$ 是线性无关的. 从最后 $n-k$ 列选择 $m-k$ 列与前 k 列一起组成一个线性独立向量组, 不妨假设选取的列是 $a_{k+1}, \cdots, a_m$, 于是 A 可以表示为 $A=[B, N]$ 形式, 其中 $B=(a_1, \cdots, a_m)$ 是满秩矩阵, 而且

$$B^{-1}b=(x_1, \cdots, x_k, 0, \cdots, 0)^{\mathrm{T}}.$$

又因为 $x_i>0,\ i=1, \cdots, k$, 所以

$$B^{-1}b \geqslant 0.$$

定理得证.

下述定理给出极方向的充要条件.

定理 1.4.2 $\bar{d} \in \mathbf{R}^n$ 为集合 (1.4.2) 的极方向当且仅当 A 能够分解为 $A=[B, N]$ 形式, 其中 B 是 $m \times n$ 阶可逆阵, N 是 $m \times (n-m)$ 阶矩阵, 使得对 N 的某列 α_j, 有 $B^{-1}\alpha_j \leqslant 0$, 并且 $\bar{d}$ 是 $d=\begin{pmatrix} -B^{-1}\alpha_j \\ e_j \end{pmatrix}$ 的正数倍, 其中 e_j 为第 j 个分量为 1, 其余为零的 $n-m$ 维单位向量.

证明 充分性. 从集合 S 的结构易见, 如果 $Ay=0, y \geqslant 0$, 则 y 是 S 的一个方向. 假如 $B^{-1}\alpha_j \leqslant 0$, 于是 $d \geqslant 0$, $Ad=0$, 此时 d 是 S 的一个方向. 下面证明 d 是一个极方向. 假设 $d=\lambda_1 d_1+\lambda_2 d_2$, 其中 $\lambda_1, \lambda_2>0$, d_1 和 d_2 是 S 的方向. 注意到 d 的 $n-m-1$ 个分量是零, 则 d_1, d_2 的相应分量也为零, 于是 d_1 和 d_2 可以写成如下形式:

$$d_1=c_1\begin{pmatrix} d_{11} \\ e_j \end{pmatrix}, \quad d_2=c_2\begin{pmatrix} d_{21} \\ e_j \end{pmatrix},$$

其中 $c_1,\ c_2>0$. 注意到 $Ad_1=Ad_2=0$, 容易验证 $d_{11}=d_{21}=-B^{-1}\alpha_j$, 这样 d_1 和 d_2 相同, 于是 d 是极方向. 因为 $\bar{d}$ 是 d 的正整数倍, $\bar{d}$ 也是一个极方向.

必要性. 假设 $\bar{d}$ 是 S 的一个极方向, 不失一般性, 假设

$$\bar{d}=(\bar{d}_1, \cdots, \bar{d}_k, 0, \cdots, \bar{d}_j, \cdots, 0)^{\mathrm{T}},$$

其中 $\bar{d}_i>0,\ i=1, \cdots, k$ 和 $i=j$. 我们要证明 $\alpha_1, \cdots, \alpha_k$ 线性无关. 用反证法, 假设 $\alpha_1, \cdots, \alpha_k$ 线性相关, 则存在一组不全为零的数 $\lambda_1, \cdots, \lambda_k$, 使得 $\sum\limits_{i=1}^{k} \lambda_i \alpha_i=0$. 令 $\lambda=(\lambda_1, \cdots, \lambda_k, 0, \cdots, 0)^{\mathrm{T}}$, 并选取 $\alpha>0$ 充分小, 使得 $d_1=\bar{d}+\alpha\lambda$ 和 $d_2=\bar{d}-\alpha\lambda$ 都非负. 显然

$$\begin{aligned} Ad_1 &= A\bar{d}+\alpha A\lambda \\ &= 0+\alpha\sum_{i=1}^{k}\alpha_i\lambda_i \\ &= 0. \end{aligned}$$

类似有 $Ad_2=0$. 又因为 $d_1,d_2\geqslant 0$, 所以 d_1,d_2 是 S 的方向. 因为 $\alpha>0,\lambda\neq 0$, 所以 $d_1\neq d_2$, 但是 $\bar{d}=\frac{1}{2}d_1+\frac{1}{2}d_2$, 这与假设 $\bar{d}$ 是一个极方向矛盾, 因而 $a_1,\cdots,a_k$ 是线性独立的. 由于 A 的秩等于 m, 显然有 $k\leqslant m$, 则向量组 $\{a_i|i=k+1,\cdots,n,i\neq j\}$ 中必定存在 $m-k$ 个向量同 $a_1,\cdots,a_k$ 一起形成线性独立的向量组. 不失一般性, 假设这些向量是 $a_1,\cdots,a_m$, 用 B 表示 $(a_1,\cdots,a_m)$. 注意到 B 是可逆的, 于是

$$0=A\bar{d}=B\widehat{d}+\alpha_j\bar{d}_j,$$

其中 $\widehat{d}$ 是 $\bar{d}$ 的前 m 个分量形成的向量, 所以 $\widehat{d}=-\bar{d}_jB^{-1}d_j$, 因此向量 $\bar{d}$ 有形式

$$\bar{d}=\bar{d}_j\begin{pmatrix}-B^{-1}\alpha_j\\ e_j\end{pmatrix}.$$

注意到 $\bar{d}\geqslant 0$, $\bar{d}_j>0$, 于是 $B^{-1}\alpha_j\leqslant 0$. 定理得证.

上述两个定理给出了集合 (1.4.2) 极点和极方向的刻画, 事实上, 我们很容易证明它的极点和极方向都是有限的. 极点的个数不超过

$$\begin{pmatrix}n\\ m\end{pmatrix}=\frac{n!}{m!(n-m)!},$$

极方向的个数不超过

$$(n-m)\begin{pmatrix}n\\ m\end{pmatrix}=\frac{n!}{m!(n-m-1)!}.$$

另外, 极点一定是存在的, 极方向则不然, 只有当 S 为无界时才有极方向.

1.4.2 多面体的极点和极方向表示

极点和极方向的一个重要的应用就是凸多面体总可以表示为它的极点的凸组合和极方向的非线性组合. 下述定理将给出此结果.

定理 1.4.3 设 $x_1,\cdots,x_k$ 和 $d_1,\cdots,d_l$ 分别是集合 (1.4.2) 的极点和极方向, 则 $x\in S$ 当且仅当 x 可以表示为

$$x=\sum_{i=1}^{k}\lambda_ix_i+\sum_{j=1}^{l}\mu_jd_j,$$

其中 $\lambda_i\geqslant 0,i=1,\cdots,l,\mu_j\geqslant 0,j=1,\cdots,l$, 且有 $\sum\limits_{i=1}^{k}\lambda_i=1$.

证明 构造下面的集合:

$$\Lambda=\left\{\sum_{i=1}^{k}\lambda_ix_i+\sum_{j=1}^{l}\mu_jd_j\,\middle|\,\sum_{i=1}^{k}\lambda_i=1,\lambda_i\geqslant 0,i=1,\cdots,k,\mu\geqslant 0,j=1,\cdots,l\right\},$$

易见集合 Λ 是闭凸集且非空. 显然, Λ 为集合 (1.4.2) 子集, 为证明集合 Λ 与集合 (1.4.2) 相同, 只需证明集合 (1.4.2) 为 Λ 的子集. 用反证法, 假设 z 属于集合 (1.4.2), 但不属于 Λ, 由定理 1.3.1 可知, 存在 $p \in \mathbf{R}^n$ 和常数 α, 使得对于满足

$$\sum_{i=1}^{k} \lambda_i = 1, \quad \lambda_i \geqslant 0, \quad N_j \geqslant 0$$

的 $\lambda_i \geqslant 0, i = 1, \cdots, l$ 和 $\mu_j \geqslant 0, j = 1, \cdots, l$, 有

$$p^{\mathrm{T}} z > \alpha,$$

$$p^{\mathrm{T}} \left(\sum_{i=1}^{k} \lambda_i x_i + \sum_{j=1}^{l} \mu_j d_j \right) \leqslant \alpha. \tag{1.4.5}$$

因为 μ_j 可以任意大, 式 (1.4.5) 仅当 $p^{\mathrm{T}} z \leqslant 0, j = 1, \cdots, l$ 时成立. 在式 (1.4.5) 中, 通过对一切 j, 令 $\mu_j = 0, \lambda_j = 1$ 和对 $i \neq j$ 令 $\lambda_i = 0$, 则得到 $p^{\mathrm{T}} x_i \leqslant \alpha$. 因为 $p^{\mathrm{T}} z > \alpha$, 对一切 i 有 $p^{\mathrm{T}} z > p^{\mathrm{T}} x_i$, 综合可得, 存在一个非零向量 p, 使得

$$p^{\mathrm{T}} z > p^{\mathrm{T}} x_i, \quad i = 1, \cdots, k, \tag{1.4.6}$$

$$p^{\mathrm{T}} d_j, \quad j = 1, \cdots, l. \tag{1.4.7}$$

考虑如下定义的极点 $\bar{x}$:

$$p^{\mathrm{T}} \bar{x} = \max_{1 \leqslant j \leqslant k} p^{\mathrm{T}} x_j. \tag{1.4.8}$$

因为 $\bar{x}$ 是极点, 由定理 1.4.1, $\bar{x} = \begin{pmatrix} B^{-1} b \\ 0 \end{pmatrix}$, 其中 $A = (B, N)$, $B^{-1} b \geqslant 0$. 不失一般性, 假定 $B^{-1} b > 0$. 因 z 为集合 (1.4.2) 的点, 于是

$$Az = b, \quad z \geqslant 0,$$

所以

$$B z_B + N z_N = b,$$

进而有

$$z_B = b^{-1} b - B^{-1} N z_N,$$

其中 z 被分解为 (z_B, z_N). 从式 (1.4.6) 得

$$p^{\mathrm{T}} z - p^{\mathrm{T}} \bar{x} > 0,$$

又 p^{T} 被分解为 (p_B, p_N), 得到

$$\begin{aligned}0 &< p^{\mathrm{T}}z - p^{\mathrm{T}}\bar{z} \\ &= p_B^{\mathrm{T}}(B^{-1}b - B^{-1}Nz_N) + p_N^{\mathrm{T}}z_N - p_B^{\mathrm{T}}B^{-1}b \\ &= (p_N^{\mathrm{T}} - p_B^{\mathrm{T}}B^{-1}N)z_N.\end{aligned} \tag{1.4.9}$$

因为 $z \geqslant 0$, 式 (1.4.9) 得到一个分量 $j \geqslant m+1$, 使得

$$z_j > 0, \quad p_j - p_B^{\mathrm{T}}B^{-1}\alpha_j > 0.$$

首先证明 $y_i^{\mathrm{T}}B^{-1}\alpha_j$ 不小于 0, 用反证法, 假设 $y_j \leqslant 0$. 考虑向量

$$d_j = \begin{pmatrix} -y_j \\ e_j \end{pmatrix},$$

其中 e_i 是第 i 个分量为 1 的单位向量. 由定理 1.4.2, d_j 是集合 (1.4.2) 的一个极方向. 由式 (1.4.7) 得

$$p^{\mathrm{T}}d_j \leqslant 0,$$

即

$$-p_B^{\mathrm{T}}B^{-1}\alpha_j + p_j \leqslant 0,$$

这同假设

$$p_j - p_B^{\mathrm{T}}B^{-1}\alpha_j > 0$$

相矛盾. 因此, y_j 不小于 0. 构造下面的向量:

$$x = \begin{pmatrix} \bar{b} \\ 0 \end{pmatrix} + \lambda \begin{pmatrix} -y_j \\ e_j \end{pmatrix},$$

其中 $\bar{b}$ 由 $B^{-1}b$ 给出, λ 由下式确定

$$\lambda = \min_{1 \leqslant i \leqslant m} \left\{ \frac{\bar{b}_i}{y_{ij}} | y_{ij} > 0 \right\} = \frac{\bar{b}_r}{y_{rj}} > 0.$$

注意到, $x \geqslant 0$ 至多有 m 个正分量, 其中第 r 个分量下降到零, 而第 j 个分量给定为 λ. 向量 x 属于集合 (1.4.2), 这是因为

$$Ax = b,$$

$$(B^{-1}b - \lambda B^{-1}\alpha_j) + \lambda\alpha_j = b.$$

由于 $y_{rj}\neq 0$, 能够证明向量 $\alpha_1,\cdots,\alpha_{j-1},\alpha_{j+1},\cdots,\alpha_m,\alpha_j$ 是线性独立的. 因此, 定理 1.4.1 保证 x 是一个极点, 即 $x\in\{x_1,\cdots,x_k\}$. 此外,

$$\begin{aligned}p^{\mathrm{T}}x&=(p_B^{\mathrm{T}},p_N^{\mathrm{T}})\begin{pmatrix}\bar{b}-\lambda y_j\\ \lambda e_j\end{pmatrix}\\&=p_B^{\mathrm{T}}\bar{b}-\lambda p_B^{\mathrm{T}}y_j+\lambda p_j\\&=p_B^{\mathrm{T}}\bar{x}+\lambda(p_j-p_B^{\mathrm{T}}B^{-1}\alpha_j),\end{aligned}$$

由于

$$\lambda>0,p_j-p_B^{\mathrm{T}}B^{-1}\alpha_j>0,$$

则

$$p^{\mathrm{T}}x>p^{\mathrm{T}}\bar{x}_j.$$

至此, 我们构造了一个极点 x, 使得

$$p^{\mathrm{T}}x>p^{\mathrm{T}}\bar{x}_j,$$

这同 (1.4.8) 矛盾, 这一矛盾说明 z 必属于 Λ. 定理得证.

1.5　相对内部

将 $n-1$ 维空间中一个有内部的集合放到 n 维空间中则无内部, 由此可见, 内部的概念依赖于所在空间, 不是一个独立于空间维数的概念. 为了研究凸集不依赖于所在空间维数的内部, 本节引入凸集的相对内部.

1.5.1　仿射集

在 $\mathbf{R}^2$ 中, 过原点的直线 l_1 是 $\mathbf{R}^2$ 中的子空间, 不过原点的直线 l_2 则不是子空间, 但它与 l_1 平行, 是 l_1 的平移.

定义 1.5.1　设 $S\subset\mathbf{R}^n$, 对于任意 x, $y\in S$ 和常数 λ, 均有 $(1-\lambda)x+\lambda y\in S$, 则称 S 是 $\mathbf{R}^n$ 中的仿射集. 仿射集有时也称为仿射流形、仿射族、线性流形等.

从定义 1.5.1 可以看出, 通过仿射集 S 中任意两点的直线仍然包含在 S 中.

例 1.5.1　$\mathbf{R}^n$ 空间的空集、全空间、单点集、直线和平面都是仿射集.

定义 1.5.1 表示, 仿射集中任意两个点的仿射组合仍然属于该仿射集中, 下面的结果显然成立.

定理 1.5.1　设 $S\subset\mathbf{R}^n$, 则 S 是 $\mathbf{R}^n$ 中的仿射集的充要条件是 S 的元素的任意仿射组合仍然属于 M.

容易证明, 仿射集的交仍是仿射集. 所以对于 $\mathbf{R}^n$ 中的任意集合 S, 存在着包含 S 的最小仿射集, 这个仿射集就是包含 S 的全体仿射集的交, 称这个仿射集是由 S 张成的仿射集, 或由 S 生成的仿射集, 或 S 的仿射包, 用 aff$S$ 表示.

定义 1.5.2 设 $x_1, \cdots, x_p$ 是 $\mathbf{R}^n$ 中的向量, 常数 $\lambda_1, \cdots, \lambda_p$ 满足 $\lambda_1 + \cdots + \lambda_p = 1$, 则 $x = \lambda_1 x_1 + \cdots + \lambda_p x_p$ 称向量组 $x_1, \cdots, x_p$ 的仿射组合.

命题 1.5.1 设 $S \subset \mathbf{R}^n$, 则 S 的仿射包 aff$S$ 是 S 中的点所有仿射组合构成的集合.

定义 1.5.3 设 $x_1, \cdots, x_p$ 是 $\mathbf{R}^n$ 中的点, 常数 $\lambda_1, \cdots, \lambda_p$ 满足 $\lambda_1 + \cdots + \lambda_p = 1$, 如果仅当 $\lambda_1 = \cdots = \lambda_p = 0$ 时, 有

$$\lambda_1 x_1 + \cdots + \lambda_p x_p = 0,$$

则称向量组 $x_1, \cdots, x_p$ 仿射无关, 不是仿射无关的向量组称为仿射相关.

向量组的仿射无关性等价于这个向量组的任一向量都不是其他向量的仿射组合. 如果某一个向量 x 表示成向量组 $x_1, \cdots, x_p$ 仿射组合的形式, 即 $x = \lambda_1 x_1 + \cdots + \lambda_p x_p$, 其中 λ_i, $i = 1, \cdots, p$ 为常数且有 $\lambda_1 + \cdots + \lambda_p = 1$, 则当且仅当 $x_1, \cdots, x_p$ 仿射无关时, 系数 $\lambda_1, \cdots, \lambda_p$ 可以唯一确定.

定理 1.5.2 $\mathbf{R}^n$ 的子空间是包含原点的仿射集, 反之包含原点的仿射集是子空间.

证明 每一个子空间包含原点, 且对加法和数乘运算封闭, 因而是仿射集. 反之, 设 S 是包含原点的仿射集, 对于任意 $x \in S$ 和常数 λ, 由定义 1.5.1, 有

$$\lambda x = (1 - \lambda) 0 + \lambda x \in S,$$

即 S 对数乘运算封闭. 对于任意 $x, y \in S$, 有

$$\frac{1}{2}(x + y) = \frac{1}{2} x + \left(1 - \frac{1}{2}\right) y \in S,$$

故

$$x + y = 2 \left(\frac{1}{2}(x + y)\right) \in S,$$

即 S 对加法运算封闭, 所以 S 是子空间. 定理得证.

定义 1.5.4 设 $S \subset \mathbf{R}^n$, $a \in \mathbf{R}^n$, 集合 $M + a = \{x + a | x \in S\}$ 称为 S 平移 a 的集合.

1.5.2 相对内部的定义与性质

定义 1.5.5 设 $S \subset \mathbf{R}^n$, $x \in C$, 如果存在一个 $\varepsilon > 0$, 使 $x + B(0, \varepsilon) \subset S$, 则称 x 是 S 的内点. S 的全体内点的集合称为 S 的内部, 用 int$S$ 表示.

易见

$$\text{int}S = \{x|\exists\varepsilon > 0, x + B(0,\varepsilon) \subset S\}.$$

对于一般的凸集 S, 并不能保证 $\text{int}S \neq \varnothing$. 例如在 $\mathbf{R}^3$ 中, 三角形是没有内点的, 但是, 在由这个三角形张成的二维仿射集 (仿射包) 中, 它确实包含有内点. 下面将要证明, 总可以在某种意义下将一个非空凸集嵌入到 $\mathbf{R}^n$ 的子空间中去, 使这个凸集关于这个子空间具有内点.

定理 1.5.3 设 S 是 $\mathbf{R}^n$ 中的非空凸集, 则 S 或者有内点, 或者 S 包含在一个维数较低的仿射集之中.

证明 设 $x_0 \in S$, 研究形如 $x - x_0$ 的向量, 其中 $x \in S$. 由线性代数的有关定理知, 存在 $r \leqslant n$ 个上述形式的线性无关向量:

$$x_1 - x_0, \cdots, x_r - x_0.$$

分两种情形讨论.

(1) $r=n$. 这时 n 个向量 $x_1-x_0,\cdots,x_n-x_0$ 线性无关, 单纯形 $S^n(x_0,x_1,\cdots,x_n) \subset S$, 故只要证明 $S^n(x_0,x_1,\cdots,x_n)$ 具有内点即可. 下面证明:

$$x = \lambda_0 x_0 + \lambda_1 x_1 + \cdots + \lambda_n x_n \in \text{int}S^n(x_0, x_1, \cdots, x_n),$$

其中

$$\lambda_i > 0, \quad i = 0, 1, \cdots, n, \quad \sum_{i=0}^{n} \lambda_i = 1.$$

研究关于 $\lambda_i, i = 0, 1, \cdots, n$ 的线性方程组:

$$x - x_0 = \sum_{i=1}^{n} \lambda_i (x_i - x_0).$$

因为 $x_1-x_0,\cdots,x_n-x_0$ 线性无关, 这个方程组具有连续依赖于 x 的唯一解 $\lambda_i(x), i = 1,\cdots,n$. 特别取 x 的值为

$$\bar{x} = \bar{\lambda}_0 x_0 + \bar{\lambda}_1 x_1 + \cdots + \bar{\lambda}_n x_n,$$

故有

$$\lambda_i(\bar{x}) = \bar{\lambda}_i > 0, \quad i = 1, \cdots, n,$$

$$\lambda_0(\bar{x}) = 1 - \sum_{i=1}^{n} \bar{\lambda}_i > 0.$$

所以, 对 $\bar{x}$ 的某个邻域内的全体 x, 有 $\lambda_i(x) > 0$, $i = 0, 1, \cdots, n$, 且

$$\lambda_0(x) = 1 - \sum_{i=1}^{n} \lambda_i(x) > 0.$$

故对 $\bar{x}$ 的这个邻域内的全体 x, 有

$$x = \sum_{i=0}^{n} \lambda_i(x) x_i \in S^n(x_0, x_1, \cdots, x_n).$$

这表示形如上述的 x 是 $S^n(x_0, x_1, \cdots, x_n)$ 的内点, 定理的第一部分结论成立.

(2) $r < n$. 研究由 $x_i - x_0, i = 1, \cdots, r$ 张成的子空间 X°. 由 $x_i - x_0$ 的选取可知, $S - x_0 \subset X^\circ$, 即 $S \subset x_0 + X^\circ$. 这里 $x_0 + X^\circ$ 是 r 维的仿射集, 即 S 包含在一个 $r < n$ 的仿射集中, 第二部分结论成立. 定理得证.

定义 1.5.6 设 $S \subset \mathbf{R}^n$ 是凸集, $x \in S$, 如果存在一个 $\varepsilon > 0$, 使得

$$(x + B(0, \varepsilon)) \bigcap \operatorname{aff} S \subset S,$$

则称 x 是 S 的相对内点, S 的全体相对内点的集合称为 S 的相对内部, 用 $\operatorname{ri}S$ 表示.

定理 1.5.4 设 $S \subset \mathbf{R}^n$ 是非空凸集, 则 $\operatorname{ri}S$ 非空.

证明 设 $\dim S = r$, 则在 S 中存在 $r+1$ 个仿射无关的向量 $x_0, x_1, \cdots, x_r$. 令

$$A = \operatorname{co}\{x_0, x_1, \cdots, x_r\},$$

则 A 是 r 维单纯形, $A \subset S$. 但从定理 1.5.2 的证明可以看出, A 具有相对于 $\operatorname{aff} A$ 的非空相对内部. 又因为 $\operatorname{aff} A \subset \operatorname{aff} S$, 且 $\dim(\operatorname{aff} A) = r = \dim(\operatorname{aff} S)$, 所以有 $\operatorname{aff} A = \operatorname{aff} S$. 这表示 A 具有相对于 $\operatorname{aff} S$ 的非空相对内部. 于是, 由 A 是 S 的子集可知, S 具有相对于 $\operatorname{aff} S$ 的非空相对内部, 即 $\operatorname{ri}S$ 非空. 定理得证.

定理 1.5.5 设 $S \subset \mathbf{R}^n$ 是凸集, $x \in \operatorname{ri}S$, $y \in \operatorname{cl}S$, 则 $(1-\lambda)x + \lambda y \in \operatorname{ri}S$, $0 \leqslant \lambda < 1$.

证明 不妨设 $\dim S = n$, 故 $\operatorname{ri}S = \operatorname{int}S$. 设 $0 \leqslant \lambda < 1$, 只要证明, 对于某个 $\varepsilon > 0$, 有

$$(1-\lambda)x + \lambda y + B(0, \varepsilon) \subset S \tag{1.5.1}$$

即可. 由 $y \in \operatorname{cl}S$, 故对于每一个 $\varepsilon > 0$, $y \in S + B(0, \varepsilon)$. 于是, 对于每一个 $\varepsilon > 0$,

$$\begin{aligned}(1-\lambda)x + \lambda y + \varepsilon B &\subset (1-\lambda)x + \lambda(S + B(0,\varepsilon)) + B(0,1) \\ &= (1-\lambda)(x + \varepsilon(1+\lambda)(1-\lambda)^{-1}B(0,1)) + \lambda S.\end{aligned}$$

但 $x \in \operatorname{int}S$, 所以 ε 充分小时,

$$x + \varepsilon(1+\lambda)(1-\lambda)^{-1}B(0,1) \subset S,$$

则有

$$(1-\lambda)x + \lambda y + \varepsilon B \subset (1-\lambda)S + \lambda S = S.$$

故式 (1.5.1) 成立. 定理得证.

定理 1.5.4 是一个很有用的工具, 它在下面几个定理的证明中都起着决定性的作用.

定理 1.5.6　设 $S \subset \mathbf{R}^n$ 是非空凸集, 则 $z \in \mathrm{ri}S$ 的充分必要条件是对于每一个 $x \in S$, 存在 $\mu > 1$, 使得

$$(1-\mu)x + \mu z \in S.$$

证明　定理的条件表示 S 中以 z 作为端点的每一条线段可以稍微向外延长而不超出 S. 如果 $z \in \mathrm{ri}S$, 由定义 1.5.4, 所述条件成立. 另一方面, 设条件满足, 由定理 1.5.4, $\mathrm{ri}S$ 非空, $x \in \mathrm{ri}S$. 设

$$y = (1-\mu)x + \mu z \in S, \quad \mu > 1,$$

则有

$$z = (1-\lambda)x + \lambda y, \quad 0 < \lambda = \mu^{-1} < 1.$$

由定理 1.5.4, $z \in \mathrm{ri}S$. 定理得证.

容易证明, 当 S 是凸集时, $\mathrm{ri}S$ 也是凸集.

对于 $\mathbf{R}^n$ 中的任意集合 S, $\mathrm{cl}(\mathrm{cl}S) = \mathrm{cl}S$, $\mathrm{ri}(\mathrm{ri}S) = \mathrm{ri}S$ 总是成立的. 如果 S 是凸集, 则还有下述结果.

定理 1.5.7　设 S 是 $\mathbf{R}^n$ 中的凸集, 则有

(1) $\mathrm{cl}S = \mathrm{cl}(\mathrm{cl}S) = \mathrm{cl}(\mathrm{ri}S)$,

(2) $\mathrm{ri}S = \mathrm{ri}(\mathrm{cl}S) = \mathrm{ri}(\mathrm{ri}S)$,

(3) $\mathrm{aff}\, S = \mathrm{aff}(\mathrm{cl}\, S) = \mathrm{aff}(\mathrm{ri}\, S)$,

(4) $\dim S = \dim(\mathrm{cl}\, S) = \dim(\mathrm{ri}\, S)$.

1.6　切锥与法锥

切锥是光滑问题在切平面的推广, 可用来在一点附近逼近一个集合. 在非光滑分析中, 有各种各样的切锥, 本节主要介绍最常见的称为 Bouligand 切锥的一种切锥. 法锥则是光滑问题在法方向的推广.

定义 1.6.1　设 $S \subset \mathbf{R}^n$ 非空, 集合 S 在点 $x \in S$ 处的切锥 $T_S(x)$ 定义如下:

$$T_S(x) = \{d \in \mathbf{R}^n | x_k = x + t_k d_k \in S,\ d_k \to d,\ t_k \to 0^+\},$$

这里 $d \in T_S(x)$ 称为切方向.

上述定义的切锥 $T_S(x)$ 也称为 Bouligand 切锥. 当 S 边界为光滑曲面时, 切锥退化为通常的切平面. 由定义不难验证, 切锥是闭锥.

Bouligand 切锥还有其他的表示形式, 例如, $T_S(x)$ 可以等价的表示为

$$T_S(x) = \left\{ d \in \mathbf{R}^n \middle| \liminf_{t \to 0^+} \frac{d_s(x+td)}{t} = 0 \right\},$$

其中 $d_S(y)$ 为点 y 到集合 S 的距离.

例 1.6.1 设 $c_i(x)$ $i = 1, \cdots, m$ 是 $\mathbf{R}^n$ 上的连续可微函数, 它们的梯度 $\nabla c_1(x), \cdots, \nabla c_m(x)$ 是线性无关的, 则集合

$$S = \{x \in \mathbf{R}^n | c_i(x) = 0,\ i = 1, \cdots, m\}$$

在点 $x \in S$ 的切锥是下述子空间

$$T_S(x) = \{d \in \mathbf{R}^n | \nabla c_i(x)^{\mathrm{T}} d = 0,\ i = 1, \cdots, m\}.$$

例 1.6.2 设 $c(x)$ 是 $\mathbf{R}^n$ 上的连续可微函数, 集合

$$S = \{x \in \mathbf{R}^n | c(x) \leqslant 0\}$$

在点 $x \in S$ 的切锥为半空间

$$T_S(x) = \{d \in \mathbf{R}^n | \nabla c(x)^{\mathrm{T}} d \leqslant 0\}.$$

定义 1.6.2 设 $S \subset \mathbf{R}^n$ 非空, 集合 S 在点 $x \in S$ 处的法锥 $N_S(x)$ 定义如下:

$$N_S(x) = \{d \in \mathbf{R}^n | d^{\mathrm{T}}(y - x) \leqslant 0\},$$

这里 $d \in T_S(x)$ 称为法方向.

当 S 边界为光滑曲面时, 法锥退化为通常的法方向. 由定义不难验证, 法锥是闭凸锥.

第2章 凸 函 数

凸函数和凸集一样, 是非光滑分析与优化中最重要的内容. 本章讨论凸函数的重要性质, 如无特别声明, 本章讨论的函数是在广义的实数轴上取值的, 即可以取值 $-\infty$ 和 $+\infty$.

2.1 凸函数基本性质

定义 2.1.1 设 S 为 $\mathbf{R}^n$ 上的非空凸紧集, $f(x)$ 为定义于 S 到 $\mathbf{R}\bigcup\{+\infty\}$ 上的函数, 如果 $f(x)$ 不恒等于 $+\infty$, 且对任意 $x,y\in S$, $0\leqslant\lambda\leqslant 1$, 有

$$f(\lambda x+(1-\lambda)y)\leqslant\lambda f(x)+(1-\lambda)f(y), \tag{2.1.1}$$

则称 $f(x)$ 为 S 上的凸函数. 如果当 $x\neq y$ 时, 式 (2.1.1) 中严格不等式成立, 则称 $f(x)$ 为严格凸函数. 不取值 $-\infty$ 且不恒等于 $+\infty$ 的凸函数称为正常凸函数, 不是正常的凸函数称为非正常凸函数.

例 2.1.1 利用定义 2.1.1 可以验证下面的初等函数是 $\mathbf{R}$ 上的凸函数.

(1) $f(x)=\mathrm{e}^x$;

(2) $f(x)=|x|$;

(3) $f(x)=\begin{cases}-\ln x, & x>0,\\ +\infty, & x\leqslant 0.\end{cases}$

定义 2.1.2 $\mathbf{R}^n$ 上的凸函数的有效域定义如下:

$$\operatorname{dom} f=\{x\in\mathbf{R}^n|f(x)<+\infty\}.$$

容易验证, 凸函数的有效域一定是凸集, 反过来, 有效域为凸集的函数不一定是凸函数.

定义于凸集 $S\subset\mathbf{R}^n$ 上的凸函数都可扩展到 $\mathbf{R}^n$ 上的凸函数, 例如, 定义 $\bar{f}(x)$ 如下:

$$\bar{f}(x)=\begin{cases}f(x), & x\in S,\\ +\infty, & x\notin S,\end{cases}$$

易见, $\bar{f}(x)$ 为定义于 $\mathbf{R}^n$ 上的凸函数. 因此, 一般来讲我们考虑的凸函数都是定义于全空间的.

定义 2.1.3 设 $f(x)$ 为 $\mathbf{R}^n$ 上的函数, 如果存在常数 $c>0$, 使得对任意 $x,y\in\mathbf{R}^n$ 和 $0\leqslant\lambda\leqslant 1$, 有

$$f(\lambda x+(1-\lambda)y)\leqslant \lambda f(x)+(1-\lambda)f(y)-\frac{1}{2}c\lambda(1-\lambda)\|x-y\|^2, \qquad (2.1.1)$$

则称 $f(x)$ 为强凸函数 (关于常数 c).

命题 2.1.1 设 $f(x)$ 为定义于 $\mathbf{R}^n$ 上的函数, $f(x)$ 是强凸函数 (关于常数 c) 的充要条件是 $f(x)-\frac{1}{2}c\|x\|^2$ 为凸函数.

证明 根据凸函数的定义, $f(x)-\frac{1}{2}c\|x\|^2$ 为凸函数等价于

$$\begin{aligned}&f(\lambda x+(1-\lambda)y)-\frac{1}{2}c\|\lambda x+(1-\lambda)y\|^2\\&\leqslant \lambda f(x)+(1-\lambda)f(y)-\frac{1}{2}c(\lambda\|x\|^2+(1-\lambda)\|y\|^2).\end{aligned} \qquad (2.1.2)$$

经过整理可知, 式 (2.1.1) 与式 (2.1.2) 等价. 命题得证.

如果 $-f(x)$ 是凸函数, 则称 $f(x)$ 是凹函数. 既是凸函数又是凹函数的函数称为仿射函数, 线性函数既是凸函数也是凹函数.

下面引入函数上图概念, 它在非光滑分析中广泛应用, 是研究和处理一些非光滑现象的一个有效工具.

定义 2.1.4 设 $f(x)$ 为定义于 $\mathbf{R}^n$ 上的函数, $f(x)$ 的上图 Epi$f$ 定义如下:

$$\text{Epi}\, f=\{(x,\alpha)\in\mathbf{R}^{n+1}|f(x)\leqslant\alpha\},$$

如果上式中 “$\leqslant$” 用 “$<$” 代替, 则 Epi f 称为严格上图.

上图和函数本身是可以相互确定的, 给定了函数的上图, 可由下式确定函数本身:

$$f(x)=\inf\{\mu|(x,\mu)\in\text{Epi}f\},$$

此处约定 $\inf\varnothing=+\infty$.

下述定理建立了函数凸性与其上图凸性的关系.

定理 2.1.1 设 $f(x)$ 为 $\mathbf{R}^n$ 上的函数, 则 $f(x)$ 为凸函数的充要条件是它的上图为 $\mathbf{R}^{n+1}$ 中的凸集.

证明 必要性. 假设 $f(x)$ 为凸函数. 记

$$z_1=(x_1,y_1)\in\text{Epi}\, f,\quad z_2=(x_2,y_2)\in\text{Epi}\, f,$$

其中 $x_1,x_2\in\mathbf{R}^n$, $y_1,y_2\in\mathbf{R}$, 对任意 $0\leqslant\lambda\leqslant 1$, 则有

$$\lambda z_1+(1-\lambda)z_2=(\lambda x_1+(1-\lambda)x_2,\ \lambda y_1+(1-\lambda)y_2).$$

由于 $f(x)$ 为凸函数及 $z_1, z_2 \in \text{Epi}\, f$, 则有

$$\begin{aligned} f(\lambda x_1 + (1-\lambda)x_2) &\leqslant \lambda f(x_1) + (1-\lambda)f(x_2) \\ &\leqslant \lambda y_1 + (1-\lambda)y_2, \end{aligned}$$

根据上图的定义有

$$(\lambda x_1 + (1-\lambda)x_2,\ \lambda y_1 + (1-\lambda)y_2) \in \text{Epi}\, f,$$

于是 $\text{Epi}\, f$ 是 $\mathbf{R}^{n+1}$ 上的凸集.

充分性. 假设 $\text{Epi}\, f$ 是 $\mathbf{R}^{n+1}$ 中的凸集. 设 $x_1, x_2 \in \mathbf{R}^n$, $0 \leqslant \lambda \leqslant 1$, 由

$$(x_1, f(x_1)) \in \text{Epi}\, f, \quad (x_2, f(x_2)) \in \text{Epi}\, f$$

及 $\text{Epi}\, f$ 的凸性, 则有

$$\lambda(x_1, f(x_1)) + (1-\lambda)(x_2, f(x_2)) \in \text{Epi}\, f,$$

即

$$(\lambda x_1 + (1-\lambda)x_2,\ \lambda f(x_1) + (1-\lambda)f(x_2)) \in \text{Epi}\, f.$$

由上图定义, 则有

$$f(\lambda x_1 + (1-\lambda)x_2) \leqslant \lambda f(x_1) + (1-\lambda)f(x_2),$$

这说明 $f(x)$ 为 $\mathbf{R}^n$ 上的凸函数. 命题得证.

定义 2.1.5　设 $f(x)$ 为定义于 $\mathbf{R}^n$ 上的函数, 下述集合:

$$L_\alpha(f) = \{x \in \mathbf{R}^n | f(x) \leqslant \alpha\}$$

称为函数 $f(x)$ 的水平集, 其中 α 为一给定常数.

容易证明, 凸函数的所有水平集是凸集, 反之则不成立, 即使一个函数的所有水平集都为凸集, 但该函数也不一定是凸函数. 所有水平集为凸集的函数称为拟凸函数. 拟凸函数是一类重要的广义凸函数, 它在最优性条件, 特别是 Kunh-Tucker 充分性条件建立中起到重要作用.

利用上图概念, 我们得到凸函数的下述 Jensen 不等式.

命题 2.1.2 (Jensen 不等式)　设 $f(x)$ 为 $\mathbf{R}^n$ 上的凸函数,

$$x_1, \cdots, x_m \in \mathbf{R}^n, \quad \lambda_1, \cdots, \lambda_m \geqslant 0$$

且有

$$\sum_{i=1}^{m} \lambda_i = 1,$$

则下述不等式成立:

$$f(\lambda_1 x_1 + \cdots + \lambda_m x_m) \leqslant \lambda_1 f(x_1) + \cdots + \lambda_m f(x_m). \tag{2.1.3}$$

证明 由于

$$(x_i, f(x_i)) \in \text{Epi}\, f, \quad i = 1, \cdots, m,$$

根据定理 2.1.1, Epi f 为 $\mathbf{R}^{n+1}$ 上的凸集, 所以 $(x_i, f(x_i))$, $i = 1, \cdots, m$ 的凸组合也在 Epi f 中, 即

$$\sum_{i=1}^{m} \lambda_i (x_i, f_i(x)) \in \text{Epi}\, f,$$

或

$$\left(\sum_{i=1}^{m} \lambda_i x_i, \sum_{i=1}^{m} \lambda_i f_i(x) \right) \in \text{Epi}\, f,$$

故式 (2.1.3) 成立. 命题得证.

定义 2.1.6 设 $f(x)$ 为 $\mathbf{R}^n$ 上函数, 如果对任意 $x \in \mathbf{R}^n$, $0 < \lambda < +\infty$, 有

$$f(\lambda x) = \lambda f(x),$$

则称 $f(x)$ 为正齐次函数.

例 2.1.2 $f(x) = |x|$ 为 $\mathbf{R}$ 上的正齐次函数.

命题 2.1.3 设 $f(x)$ 为 $\mathbf{R}^n$ 上的正齐次函数, 则 $f(x)$ 的上图为 $\mathbf{R}^{n+1}$ 空间上的锥.

证明 设 $\lambda > 0$, 利用 $f(x)$ 的正齐次性, 直接推导得

$$\begin{aligned} \lambda \text{Epi}\, f &= \lambda \{(x, \mu) | f(x) \leqslant \mu\} \\ &= \{(\lambda x, \lambda \mu) | f(x) \leqslant \mu\} \\ &= \{(\lambda x, \lambda \mu) | f(\lambda x) \leqslant \lambda \mu\} \\ &= \{(y, k) | f(y) \leqslant k\} \\ &= \text{Epi}\, f, \end{aligned}$$

这说明 Epi f 是一个锥. 命题得证.

命题 2.1.4 设 $f(x)$ 为 $\mathbf{R}^n$ 上的实值正齐次函数, 则 $f(x)$ 是凸函数的充要条件是对任意 $x, y \in \mathbf{R}^n$, 有

$$f(x + y) \leqslant f(x) + f(y). \tag{2.1.4}$$

证明　必要性. 假设 $f(x)$ 是凸函数, 根据 $f(x)$ 的正齐次和凸性得

$$\begin{aligned} f(x+y) &= 2f\left(\frac{1}{2}(x+y)\right) \\ &\leqslant 2\left(\frac{1}{2}f(x)+\frac{1}{2}f(y)\right) \\ &= f(x)+f(y), \end{aligned}$$

即式 (2.1.4) 成立.

充分性. 设式 (2.1.4) 成立, 对任意 $0 \leqslant \lambda \leqslant 1$, 由 $f(x)$ 正齐次性得

$$\begin{aligned} f((1-\lambda)x+\lambda y) &\leqslant f((1-\lambda)x)+f(\lambda y) \\ &= (1-\lambda)f(x)+\lambda f(y), \end{aligned}$$

故 $f(x)$ 是凸函数. 命题得证.

推论 2.1.1　设 $f(x)$ 为 $\mathbf{R}^n$ 上的正齐次凸函数, $\lambda_i \geqslant 0, i=1,\cdots,m$, 则有

$$f(\lambda_1 x_1+\cdots+\lambda_m x_m) \leqslant \lambda_1 f(x_1)+\cdots+\lambda_m f(x_m).$$

推论 2.1.2　设 $f(x)$ 为 $\mathbf{R}^n$ 上的正齐次凸函数, 则对任意 $x \in \mathbf{R}^n$, 有 $f(-x) \geqslant -f(x)$.

例 2.1.3　设 $S \subset \mathbf{R}^n$ 为非空凸集, 函数

$$\mu_S(x) = \inf\{\lambda | \lambda \geqslant 0,\ x \in \lambda S\}$$

称为集合 S 的规范函数, 或 Minkowski 函数. 容易验证, 规范函数是凸函数.

定义 2.1.7　满足条件:

$$f(\lambda x) = \lambda f(x), \quad \lambda \geqslant 0,$$

$$f(x+y) \leqslant f(x)+f(y)$$

的函数称为次线性函数.

定义 2.1.8　设 $S \subset \mathbf{R}^n$ 为非空凸紧集, 函数

$$\sigma_S^*(x) = \max_{s \in S} s^{\mathrm{T}} x \tag{2.1.5}$$

称为 S 的支撑函数, 支撑函数有时也记为 $P_S(x)$ 或 $\sigma^*(x|\,S)$.

由式 (2.1.5) 不难看出

$$\begin{aligned} \sigma_S^*(\lambda x) &= \lambda \max_{s \in S} s^{\mathrm{T}} x \\ &= \lambda \sigma(x), \quad \lambda \geqslant 0, \\ \sigma_S^*(x+y) &\leqslant \max_{s \in S} s^{\mathrm{T}} x + \max_{s \in S} s^{\mathrm{T}} y \\ &= \sigma_S(x)+\sigma_S(y), \end{aligned}$$

故支撑函数 $\sigma_S^*(x)$ 是正齐次函数.

例 2.1.4 设 $S \subset \mathbf{R}^n$ 为非空凸集, 函数

$$d_S = \inf_{y \in S} \|y - x\|$$

称为距离函数. 以下证明距离函数是凸函数. 设 $x, z \in \mathbf{R}^n$, 在集合 S 中, 选取两组序列 $\{y_k\}$ 和 $\{y_k'\}$, 使得

$$\|y_k - x\| \to d_S(x), \quad k \to +\infty,$$

$$\|y_k' - z\| \to d_S(z), \quad k \to +\infty.$$

对任意 $0 \leqslant \lambda \leqslant 1$, 由 S 的凸性有

$$\lambda y_k + (1-\lambda) y_k' \in S,$$

于是

$$\begin{aligned} d_S(\lambda x + (1-\lambda) z) &\leqslant \|\lambda y_k + (1-\lambda) y_k' - \lambda x - (1-\lambda) y\| \\ &\leqslant \lambda \|y_k - x\| + (1-\lambda) \|y_k' - y\|. \end{aligned}$$

对上式右端关于 $k \to \infty$ 取极限得

$$d_S(\lambda x + (1-\lambda) y) \leqslant \lambda d_S(x) + (1-\lambda) d_S(z),$$

这就证明了距离函数 $d_S(x)$ 的凸性.

2.2 凸函数代数运算

本节讨论凸函数的一些代数运算, 主要是一些保持凸性的运算.

2.2.1 复合运算

定理 2.2.1 设 $f(x)$ 为 $\mathbf{R}^n$ 上的实值凸函数, $\varphi(z)$ 是 $\mathbf{R}$ 上的非减凸函数, 则复合函数 $h(x) = \varphi(f(x))$ 是 $\mathbf{R}^n$ 上的凸函数.

证明 对任意 $x, y \in \mathbf{R}^n$, $0 \leqslant \lambda \leqslant 1$, 根据 $f(x)$ 的凸性有

$$f(\lambda x + (1-\lambda) y) \leqslant \lambda f(x) + (1-\lambda) f(y). \tag{2.2.1}$$

根据 $\varphi(z)$ 的单调性和凸性, 将式 (2.2.1) 应用于函数 $\varphi(z)$ 直接计算得

$$\begin{aligned} h(\lambda x + (1-\lambda) y) &= \varphi(f(\lambda x + (1-\lambda) y)) \\ &\leqslant \varphi(\lambda f(x) + (1-\lambda) f(y)) \\ &\leqslant \lambda \varphi(f(x)) + (1-\lambda) \varphi(f(y)) \\ &= \lambda h(x) + (1-\lambda) h(x), \end{aligned}$$

故 $h(x)$ 是凸函数. 定理得证.

利用定理 2.2.1 可以验证一些复合函数的凸性.

例 2.2.1 设 $f(x)$ 是 $\mathbf{R}^n$ 上的实值凸函数, 则 $h(x)=\mathrm{e}^{f(x)}$ 也是凸函数. 取 $\varphi(z)=\mathrm{e}^z$, 易见 $\varphi(z)$ 是 $\mathbf{R}$ 上的非减凸函数, 利用定理 2.2.1 即得 $h(x)$ 的凸性.

例 2.2.2 设 $f(x)$ 是非负凸函数, $p>1$, 取

$$\varphi(y)=\begin{cases} y^p, & y\geqslant 0,\\ 0, & y<0, \end{cases}$$

利用定理 2.2.1, 则知 $h(x)=\varphi(f(x))=(f(x))^p$ 是凸函数.

命题 2.2.1 设 $f(x)$ 是凸函数, 则它的非负乘积 $\lambda f(x)$ 也是凸函数.

证明 取 $\varphi(y)=\lambda y, y\in\mathbf{R},\ \lambda\geqslant 0$, 则 $\varphi(y)$ 是不减凸函数, 由定理 2.2.1, $\lambda f(x)=\varphi(f(x))$ 是凸函数. 命题得证.

2.2.2 凸函数与上图的关系

由上一节可知, $\mathbf{R}^n$ 上的凸函数都伴随着 $\mathbf{R}^{n+1}$ 中的一个上图, 并且凸函数可由它的上图来表示, 即

$$f(x)=\inf\{\mu|(x,\mu)\in \mathrm{Epi}\, f\}. \tag{2.2.2}$$

现在我们考虑它的反问题: $\mathbf{R}^{n+1}$ 中的凸集是否可以对应 $\mathbf{R}^n$ 中的一个凸函数呢? 下面定理给出了肯定的回答.

定理 2.2.2 设 $F\subset\mathbf{R}^{n+1}$ 为凸集, 定义函数

$$f(x)=\inf\{\mu|(x,\mu)\in F\},\quad x\in\mathbf{R}^n,$$

则 $f(x)$ 为 $\mathbf{R}^n$ 上的凸函数.

证明 设 $x,y\in\mathbf{R}^n$, 不妨假设

$$f(x)\neq+\infty, f(y)\neq+\infty,$$

于是存在 $\alpha,\beta\in\mathbf{R}$, 使得

$$f(x)<\alpha, f(y)<\beta.$$

由函数 $f(x)$ 的定义, 存在 $\mu,\gamma\in\mathbf{R}$ 满足 $\mu<\alpha,\gamma<\beta$, 使得

$$(x,\mu)\in F,(y,\gamma)\in F.$$

根据集合 F 的凸性, 则有

$$\lambda(x,\mu)+(1-\lambda)(y,\gamma)=(\lambda x+(1-\lambda)y,\lambda\mu+(1-\lambda)\gamma)\in F,$$

于是

$$f(\lambda x+(1-\lambda)y)\leqslant \lambda\mu+(1-\lambda)\gamma<\lambda\alpha+(1-\lambda)\beta,$$

这说明 $f(x)$ 为 $\mathbf{R}^n$ 上的凸函数. 定理得证.

2.2.3 卷积

给定凸函数 $f_1(x)$ 和 $f_2(x)$, 它们上图的和 Epi f_1 + Epi f_2 也是 $\mathbf{R}^{n+1}$ 中的凸集, 将它作为上图也可以导出 $\mathbf{R}^n$ 中的一个凸函数, 这个凸函数就是 $f_1(x)$ 和 $f_2(x)$ 的卷积.

定义 2.2.1 设 $f_1(x)$ 和 $f_2(x)$ 为 $\mathbf{R}^n$ 上的凸函数, $f_1(x)$ 和 $f_2(x)$ 的卷积定义如下:

$$(f_1\Box f_2)(x)=\inf\{f_1(x_1)+f_2(x_2)|x_1+x_2=x\}, \tag{2.2.3}$$

式 (2.2.3) 也可表示为

$$(f_1\Box f_2)(x)=\inf_{y\in\mathbf{R}^n}(f_1(y)+f_2(x-y)).$$

定理 2.2.3 设 $f_1(x)$ 和 $f_2(x)$ 为 $\mathbf{R}^n$ 上凸函数, 则卷积 $(f_1\Box f_2)(x)$ 也为 $\mathbf{R}^n$ 上凸函数, 且其上图为 Epi f_1 + Epi f_2.

证明 令

$$F_i=\text{Epi } f_i,\ i=1,2,\ F=F_1+F_2,$$

则 F 是 $\mathbf{R}^{n+1}$ 中的凸集. 根据集合 F 的定义, $(x,\mu)\in F$ 当且仅当存在

$$x_i\in\mathbf{R}^n,\quad \mu_i\in\mathbf{R},\quad (x_i,\mu_i)\in F,\quad i=1,2,$$

且

$$x=x_1+x_2,\quad \mu=\mu_1+\mu_2.$$

以 F 为上图构造凸函数, 根据定理 2.2.2, $\inf\{\mu|(x,\mu)\in F\}$ 为 $\mathbf{R}^n$ 上的凸函数, 而它正是 $(f_1\Box f_2)(x)$. 定理得证.

例 2.2.3 设 $S\subset\mathbf{R}^n$ 为凸集, 下述函数:

$$\delta_S(x)=\begin{cases}0, & x\in S,\\ +\infty, & x\notin S\end{cases}$$

称为集合 S 的指示函数. 不难验证, 指示 $\delta_S(x)$ 是凸函数. 利用定义, 求 $f(x)=\|x\|$ 和 $\delta(x)$ 的卷积得

$$\begin{aligned}(f\Box\delta)(x)&=\inf_{y\in\mathbf{R}^n}\{\|x-y\|+\delta(y)\}\\&=\inf_{y\in S}\|x-y\|=d_S(x),\end{aligned}$$

这里的 $d_S(x)$ 为距离函数. 这也证明了距离函数是凸函数.

例 2.2.4 设 $f_i(x)=\dfrac{1}{2}x^{\mathrm{T}}A_ix, i=1,2$, 其中 A_1,A_2 是正定矩阵, $f_1(x)$ 和 $f_2(x)$ 的卷积为

$$\begin{aligned}(f_1\Box f_2)&=\frac{1}{2}\inf_y\{y^{\mathrm{T}}A_1y+(x-y)^{\mathrm{T}}A_2(x-y))\\&=\frac{1}{2}x^{\mathrm{T}}(A_1^{-1}+A_2^{-1})^{-1}x.\end{aligned}$$

$\mathbf{R}^n$ 上 m 个凸函数 $f_1(x),\cdots,f_m(x)$ 的卷积定义如下:

$$(f_1\Box\cdots\Box f_m(x))=\inf\{f_1(x_1)+\cdots+f_m(x)|x_1+\cdots+x_m=x\}.$$

易见

$$\frac{1}{m}(f\Box\cdots\Box f)(x)=f\left(\frac{1}{m}x\right).$$

上式也说明, 如果 $f(x)$ 是凸函数, 则 $mf\left(\dfrac{1}{m}x\right)$ 也是凸函数, 这里 m 是正整数. 事实上, 可以证明对任意正数 m, $mf\left(\dfrac{1}{m}x\right)$ 都是凸函数.

2.2.4 最大值函数

考虑最大值函数

$$f(x)=\sup_{i\in I}f_i(x),\tag{2.2.4}$$

其中 $f_i(x), i\in I$ 是 $\mathbf{R}^n$ 上的凸函数, I 为任意指标集.

定理 2.2.4 式 (2.2.4) 给出的函数 $f(x)$ 是 $\mathbf{R}^n$ 上的凸函数.

证明 由于每个 $i\in I$, $f_i(x)$ 的上图 $\mathrm{Epi}\, f_i$ 是凸集, 故它们的交 $\bigcap\limits_{i\in I}\mathrm{Epi}\, f_i$ 是凸集. 根据式 (2.2.4) 有

$$\begin{aligned}\mathrm{Epi}\, f&=\{(x,\mu)|x\in\mathbf{R}^n,\mu\in\mathbf{R},f(x)\leqslant\mu\}\\&=\{(x,\mu)|x\in\mathbf{R}^n,\mu\in\mathbf{R},f_i(x)\leqslant\mu,i\in I\}\\&=\bigcap_{i\in I}\mathrm{Epi}\, f_i,\end{aligned}$$

这说明 $f(x)$ 的上图是凸集, 故 $f(x)$ 是凸函数. 定理得证.

例 2.2.5 设 $S\subset\mathbf{R}^n$ 为凸集, S 的支撑函数 $\delta_S^*(x)=\sup\limits_{y\in S}x^{\mathrm{T}}y$ 是线性函数 $x^{\mathrm{T}}y$(其中 y 固定) 的逐点取最大得到的, 由定理 2.2.4, 它是凸函数.

例 2.2.6 设 $x=(x_1,\cdots,x_n)\in\mathbf{R}^n$, 函数

$$f(x)=\max\{x_1,\cdots,x_n|x=(x_1,\cdots,x_n)\}$$

是线性函数 $f_i(x) = x^{\mathrm{T}} e_i$ 的逐点上确界, 其中 e_i 为 $\mathbf{R}^n$ 中第 i 个分量为 1 的单位向量, 所以 $f(x)$ 是凸函数.

2.2.5 函数的凸包

定义 2.2.2 设 $g(x)$ 是 $\mathbf{R}^n$ 上的函数, $F = \mathrm{co}(\mathrm{Epi}\ g)$, 函数 $f(x) = \inf\{\mu|(x,\mu) \in F\}$ 称为 $g(x)$ 的凸包, 通常记为 $\mathrm{co}g(x)$.

$g(x)$ 的凸包 $\mathrm{co}g(x)$ 是不大于 $g(x)$ 的最大凸函数, 它可以用来对一个非凸函数进行凸化. 如果 $g(x)$ 本身是凸函数, 那么它的凸包就是它自身.

根据定理 1.1.3, $(x,\mu) \in F = \mathrm{co}(\mathrm{Epi}\ g)$ 的等价条件是

$$\begin{aligned}(x,\mu) &= \lambda_1(x_1,\mu_1) + \cdots + \lambda_m(x_m,\mu_m) \\ &= (\lambda_1 x_1 + \cdots + \lambda_m x_m, \lambda_1\mu_1 + \cdots + \lambda_m\mu_m),\end{aligned}$$

其中

$$(x_i,\mu_i) \in \mathrm{Epi}\ g, \quad \lambda_i \geqslant 0, i = 1,\cdots,m, \quad \sum_{i=1}^{m} \lambda_i = 1.$$

于是

$$\begin{aligned}f(x) &= \inf\{\mu|(x,\mu) \in F\} \\ &= \inf\Bigg\{\lambda_1\mu_1 + \cdots + \lambda_m\mu_m | x = \lambda_1 x_1 + \cdots + \lambda_1 x_1, \\ &\qquad g(x_i) \leqslant \mu_i,\ \lambda_i \geqslant 0,\ i = 1,\cdots,m, \sum_{i=1}^{m}\lambda_i = 1\Bigg\} \\ &= \inf\Bigg\{\lambda_1 g(x_1) + \cdots + \lambda_m g(x_m) | x = \lambda_1 x_1 + \cdots + \lambda_1 x_1, \\ &\qquad \lambda_i \geqslant 0,\ i = 1,\cdots,m, \sum_{i=1}^{m}\lambda_i = 1\Bigg\}.\end{aligned}$$

2.2.6 共轭函数

定义 2.2.3 设 $f(x)$ 是 $\mathbf{R}^n$ 上的凸函数, 则称

$$f^*(x^*) = \sup\{x^{\mathrm{T}} x^* - f(x) | x \in \mathbf{R}^n\} \tag{2.2.5}$$

为 $f(x)$ 的共轭函数.

命题 2.2.2 设 $f(x)$ 是 $\mathbf{R}^n$ 上的凸函数, 则其共轭函数 $f^*(x^*)$ 也是凸函数.

证明 对固定的 $x \in \mathbf{R}^n$, 函数

$$h_x(x^*) = x^{\mathrm{T}} x^* - f(x)$$

是仿射函数, 根据共轭函数的定义, $f^*(x^*)$ 是函数 $h_x(x^*)$ 关于 x 的逐点上确界, 根据定理 2.2.4, $f^*(x^*)$ 是凸函数. 命题得证.

命题 2.2.3 (Fenchel 不等式)　设 $f(x)$ 是 $\mathbf{R}^n$ 上的凸函数, 则下述不等式成立:

$$f(x) + f^*(x^*) \geqslant x^{\mathrm{T}}x^*, \quad \forall x, x^* \in \mathbf{R}^n. \tag{2.2.6}$$

证明　由式 (2.2.5) 得

$$f^*(x^*) \geqslant x^{\mathrm{T}}x^* - f(x), \quad \forall x, x^* \in \mathbf{R}^n.$$

所以式 (2.2.6) 成立. 命题得证.

设 $f(x)$ 是 $\mathbf{R}^n$ 上的凸函数, 不难验证下述结论成立:

(1) 如果 $g(x) = f(x) + c$, 其中 c 为常数, 则 $g^*(x^*) = f^*(x^*) - c$;

(2) 如果 $g(x) = f(x + y)$, 则 $g^*(x^*) = f^*(x^*) - y^{\mathrm{T}}x^*$;

(3) 如果 $g(x) = f(ax)$, 其中 $a \neq 0$, 则 $g^*(x^*) = f^*\left(\dfrac{x^*}{a}\right)$;

(4) 如果 $g(x) = af(x)$, 其中 $a > 0$, 则 $g^*(x^*) = af^*\left(\dfrac{x^*}{a}\right)$.

例 2.2.7　设 $f(x) = \mathrm{e}^x, x \in \mathbf{R}$. 根据定义得

$$f^*(x^*) = \sup_x \{xx^* - \mathrm{e}^x\},$$

通过分析和计算得

$$f^*(x^*) = \begin{cases} x^* \ln x^* - x^*, & x^* > 0, \\ 0, & x^* = 0, \\ +\infty, & x^* < 0. \end{cases}$$

例 2.2.8　设 L 为 $\mathbf{R}^n$ 中的子空间, L 的指示函数 $\delta_L(x)$ 的共轭为

$$\begin{aligned} \sup_{x \in \mathbf{R}^n} \{x^{\mathrm{T}}x^* - \delta_L(x)\} &= \sup_{x \in \mathbf{R}^n} x^{\mathrm{T}}x^* \\ &= \begin{cases} 0, & x^* \in L^{\perp}, \\ +\infty, & x^* \in L \end{cases} \\ &= \delta_{L^{\perp}}(x^*), \end{aligned}$$

其中 $L^{\perp}$ 为 L 的正交补.

例 2.2.9　设 S 为 $\mathbf{R}^n$ 中的闭凸集, 则 S 的指示函数的共轭为 S 的支撑函数. 根据定义, $\delta_S(x)$ 的共轭函数是

$$\begin{aligned} (\delta_S(x))^*(x^*) &= \sup\{x^{\mathrm{T}}x^* - \delta_S(x) | x \in \mathbf{R}^n\} \\ &= \sup_{x \in S} x^{\mathrm{T}}x^* \\ &= \delta_S^*(x^*). \end{aligned}$$

2.3 凸函数的 Lipschitz 连续性

本节讨论凸函数的连续性, 特别是它的 Lipschitz 连续性.

定义 2.3.1 设 $f(x)$ 为定义于开集 $D \subset \mathbf{R}^n$ 上的函数, 如果对任意 $x \in D$, 存在常数 δ, L, 使得

$$|f(x_1) - f(x_2)| < L \|x_1 - x_2\|, \quad \forall x_1, x_2 \in B(x, \delta),$$

则称为 $f(x)$ 为 D 上的局部 Lipschitz 函数. 如果存在常数 L, 使对任意的 $x_1, x_2 \in D$, 都有

$$|f(x_1) - f(x_2)| \leqslant L \|x_1 - x_2\|,$$

则称 $f(x)$ 为 D 上的 Lipschitz 函数, 也称为全局 Lipschitz 函数.

显然, 局部 Lipschitz 函数是连续函数, 它实质是一种更强的连续函数, 全局 Lipschitz 函数一定是局部 Lipschitz 函数. 不难验证, 连续可微函数是局部 Lipschitz 函数.

为了证明凸函数的连续性, 首先证明以下两个常用的不等式.

定理 2.3.1 设 $f(x)$ 为 $\mathbf{R}$ 上的凸函数, $x_0 < x_1 < x_2 \in \operatorname{dom} f$, 则有下述不等式成立:

$$\frac{f(x_1) - f(x_0)}{x_1 - x_0} \leqslant \frac{f(x_2) - f(x_0)}{x_2 - x_0}, \tag{2.3.1}$$

$$\frac{f(x_1) - f(x_0)}{x_1 - x_0} \leqslant \frac{f(x_2) - f(x_1)}{x_2 - x_1}. \tag{2.3.2}$$

证明 令

$$\lambda = \frac{x_1 - x_0}{x_2 - x_0},$$

则

$$0 < \lambda < 1, \ 1 - \lambda = \frac{x_2 - x_1}{x_2 - x_0},$$

且

$$\lambda x_2 + (1 - \lambda) x_0 = \frac{x_1 - x_0}{x_2 - x_0} x_2 + \frac{x_2 - x_1}{x_2 - x_0} x_0 = x_1.$$

利用 $f(x)$ 的凸性有

$$\begin{aligned} f(x_1) &= f(\lambda x_2 + (1 - \lambda) x_0) \\ &\leqslant \frac{x_1 - x_0}{x_2 - x_0} f(x_2) + \frac{x_2 - x_1}{x_2 - x_0} f(x_0). \end{aligned} \tag{2.3.3}$$

将式 (2.3.3) 两边同时减去 $f(x_0)$ 得

$$f(x_1)-f(x_0)\leqslant\frac{x_1-x_0}{x_2-x_0}(f(x_2)-f(x_0)),$$

故式 (2.3.1) 成立.

将式 (2.3.3) 变形为

$$\frac{x_1-x_0}{x_2-x_0}f(x_1)+\frac{x_2-x_1}{x_2-x_0}f(x_1)\leqslant\frac{x_1-x_0}{x_2-x_0}f(x_2)+\frac{x_2-x_1}{x_2-x_0}f(x_0)$$

或

$$\frac{x_2-x_1}{x_2-x_0}(f(x_1)-f(x_0))\leqslant\frac{x_1-x_0}{x_2-x_0}(f(x_2)-f(x_1)),$$

故式 (2.3.2) 成立. 定理得证.

定理 2.3.2 设 $f(x)$ 是 $\mathbf{R}^n$ 上的凸函数, 则 $f(x)$ 在 $\mathrm{ri}(\mathrm{dom}\ f)$ 中的每个紧子集 S 上有上界.

证明 根据有限覆盖定理, 紧集 S 可以被顶点在 $\mathrm{dom}\, f$ 内的有限个单纯形覆盖, 故只要证明 $f(x)$ 在 $\mathrm{ri}(\mathrm{dom}\ f)$ 中的单纯形上有上界即可. 如果 $f(x)$ 是非正常函数, 结论自然成立, 因此假设 $f(x)$ 是正常凸函数.

以下按单纯形的维数应用归纳法. 由函数的凸性得

$$f(\lambda x+(1-\lambda)y)\leqslant\lambda f(x)+(1-\lambda)f(y),\quad 0\leqslant\lambda\leqslant1,$$

这说明 $f(x)$ 在一维单纯形 (一维单纯形为有界闭区间) 上是有上界的. 假设 $f(x)$ 在 $\mathrm{dom}\, f$ 内任意一个 $r-1$ 维单纯形上有上界. 设 $S^r(x_1,\cdots,x_{r+1})$ 为 r 维单纯形, 其中 $x_1,\cdots,x_{r+1}\in\mathrm{ri}(\mathrm{dom}\ f)$, 如果

$$x\in S^r(x_1,\cdots,x_{r+1}),\qquad x\neq x_{r+1},$$

则有

$$x=\lambda_1x_1+\cdots+\lambda_rx_r+\lambda_{r+1}x_{r+1},$$

$$\lambda_i\geqslant0,\quad i=1,\cdots,r+1,\quad \sum_{r=1}^{r+1}\lambda_i=1,\quad \lambda_{r+1}<1.$$

将上式改写成

$$x=\mu z+\lambda_{r+1}x_{r+1},$$

其中

$$\mu=1-\lambda_{r+1},\quad z=\frac{\lambda_1x_1+\cdots+\lambda_rx_r}{1-\lambda_{r+1}}.$$

由于 $x_1,\cdots,x_r$ 是仿射无关的, 所以 z 是 $\mathrm{dom}\, f$ 中 $r-1$ 维单纯形的点. 由归纳法假设, 存在与 $x_1,\cdots,x_r$ 无关的常数 N, 使得 $f(x)\leqslant N$, 故

$$f(x)\leqslant \mu f(z)+\lambda_{r+1}f(x_{r+1})\leqslant \max\{N,f(x_{r+1})\},\tag{2.3.4}$$

这说明 $f(x)$ 在 $S^r(x_1,\cdots,x_{r+1})$ 上有上界. 定理得证.

定理 2.3.3 设 $f(x)$ 是 $\mathbf{R}^n$ 上的正常凸函数, 且在 $x_0\in \mathrm{dom}\, f$ 的一个邻域内有上界, 则 $f(x)$ 在 x_0 处连续.

证明 不失一般性, 设 $x_0=0$. 于是, 存在常数 $\delta>0$ 和 N, 使得

$$f(x)\leqslant N,\quad x\in B(0,\delta).$$

考虑函数

$$g(\alpha)=f(\alpha x),\quad \alpha\in \mathbf{R},$$

由定义不难验证, $g(\alpha)$ 为 $\mathbf{R}$ 上的凸函数. 对函数 $g(\alpha)$ 利用式 (2.3.1), 其中取 $x_2=\ \alpha_1, x_0=0,\ x_1=1$, 得

$$\frac{g(1)-g(0)}{1}\leqslant \frac{g(\alpha)-g(0)}{\alpha},$$

而

$$g(1)=f(x)\leqslant N,\quad g(0)=f(0)\leqslant N,$$

所以

$$f(\alpha x)-f(0)\leqslant 2N\alpha.\tag{2.3.5}$$

在式 (2.3.2) 中, 令

$$x_0=-1,\quad x_1=0,\quad x_2=\alpha,$$

得

$$\frac{g(0)-g(-1)}{0-(-1)}\leqslant \frac{g(\alpha)-g(0)}{\alpha},$$

故

$$-2N\alpha\leqslant f(\alpha x)-f(0).\tag{2.3.6}$$

由式 (2.3.5) 和式 (2.3.6) 得

$$|f(\alpha x)-f(0)|\leqslant 2N\alpha.\tag{2.3.7}$$

任取 $\varepsilon>0$, 对于 $y\in B\left(0,\dfrac{\varepsilon\delta}{2N}\right)$, 则存在 $x\in B(0,\delta)$, 使得 $y=\dfrac{\varepsilon}{2N}x$, 由式 (2.3.7) 得

$$|f(y)-f(0)|=\left|f\left(\frac{\varepsilon}{2N}x\right)-f(0)\right|\leqslant 2N\frac{\varepsilon}{2N}=\varepsilon,$$

所以 $f(x)$ 在 $x_0=0$ 处连续. 定理得证.

定理 2.3.4　设 $f(x)$ 是 $\mathbf{R}^n$ 上的正常凸函数, 则 $f(x)$ 在 $\mathrm{ri}(\mathrm{dom}f)$ 上连续.

证明　设 $x_0\in \mathrm{ri}(\mathrm{dom}f)$, 则存在单纯形 $S^k(y_0,y_1,\cdots,y_k)$, 使得 x_0 为该单纯形的内点, 其中 $y_0,y_1,\cdots,y_k\in \mathrm{dom}f$, k=dim(dom$f$). 由定理 2.3.2, $f(x)$ 在 x_0 的一个邻域上有上界, 再由定理 2.3.3, $f(x)$ 在 x_0 处连续, 进而 x 在 $\mathrm{ri}(\mathrm{dom}f)$ 上连续. 定理得证.

定理 2.3.5　设 $f(x)$ 是 $\mathbf{R}^n$ 上的正常凸函数, S 是 $\mathrm{ri}(\mathrm{dom}f)$ 中的任意紧子集, 则 $f(x)$ 在 S 上 Lipschitz 连续.

证明　不失一般性, 假设 $\dim(\mathrm{dom}f)=n$, 其中 dim 代表维数, 则 $S\subset \mathrm{int}(\mathrm{dom}f)$. 对任意 $\varepsilon>0$, $S+B(0,\varepsilon)$ 是紧集, 由 $B(0,\varepsilon)$ 的闭性及 $S\subset \mathrm{int}(\mathrm{dom}f)$ 得

$$\bigcap_{\varepsilon>0}(S+B(0,\varepsilon))\bigcap(\mathbf{R}^n\backslash\mathrm{int}(\mathrm{dom}\,f))=\varnothing, \tag{2.3.8}$$

所以存在一个 $\varepsilon>0$, 使得

$$(S+B(0,\varepsilon))\bigcap(\mathbf{R}^n\backslash\mathrm{int}(\mathrm{dom}\,f))=\varnothing,$$

即

$$S+B(0,\varepsilon)\subset \mathrm{int}(\mathrm{dom}\,f).$$

由定理 2.3.3, $f(x)$ 在 $S+B(0,\varepsilon)$ 上连续, 故 $f(x)$ 在 $S+B(0,\varepsilon)$ 上有界, 设 N_2, N_1 分别是它的上、下界. 注意到 $x,y\in S$, $x\neq y$, 令

$$z=y+\frac{\varepsilon}{\|y-x\|}(y-x),$$

则

$$\|z\|\leqslant\|y\|+\left(\frac{\varepsilon}{\|y-x\|}\right)\|y-x\|=\|y\|+\varepsilon, \tag{2.3.9}$$

故 $z\in S+B(0,\varepsilon)$. 记

$$\lambda=\frac{\|y-x\|}{\varepsilon+\|y-x\|},$$

则

$$0<\lambda<1, y=(1-\lambda)x+\lambda z,$$

根据 $f(x)$ 的凸性有

$$\begin{aligned}f(y)&\leqslant(1-\lambda)f(x)+\lambda f(z)\\&=f(x)+\lambda(f(z)-f(x)),\end{aligned}$$

故

$$\begin{aligned} f(y) - f(x) &\leqslant \lambda(N_2 - N_1) \\ &= \frac{\|y - x\|}{(\varepsilon + \|y - x\|)}(N_2 - N_1) \\ &< \frac{N_2 - N_1}{\varepsilon}\|y - x\|. \end{aligned} \tag{2.3.10}$$

由 x, y 在 S 中的任意性可得 $f(x)$ 在 S 上 Lipschitz 连续. 定理得证.

推论 2.3.1 设 $f(x)$ 为 $\mathbf{R}^n$ 上的实值凸函数, 则 $f(x)$ 为 $\mathbf{R}^n$ 上的局部 Lipschitz 函数.

2.4 光滑凸函数的微分

本节介绍光滑凸函数微分的一些性质, 以及用它的微分来刻画函数的凸性, 当然要假设函数取有限值.

定理 2.4.1 设 $f(x)$ 为 $\mathbf{R}^n$ 上的可微函数, S 为 $\mathbf{R}^n$ 上的凸集, 则有下述结论:

(1) $f(x)$ 在 S 上是凸函数当且仅当

$$f(x) \geqslant f(x_0) + \nabla f(x_0)^{\mathrm{T}}(x - x_0), \quad \forall x_0, x \in S; \tag{2.4.1}$$

(2) $f(x)$ 在 S 上是严格凸的, 当且仅当对于 $x_0 \neq x$ 时, 式 (2.4.1) 中严格不等号成立;

(3) $f(x)$ 在 S 上是强凸 (关于系数 c) 的, 当且仅当

$$f(x) \geqslant f(x_0) + \nabla f(x_0)^{\mathrm{T}}(x - x_0) + \frac{1}{2}c\|x - x_0\|^2. \tag{2.4.2}$$

证明 (1) 假设 $f(x)$ 是 S 上的凸函数, 对任意 $x_0, x \in S$, $0 < \lambda < 1$, 有

$$f((1-\lambda)x_0 + \lambda x) - f(x_0) \leqslant \lambda(f(x) - f(x_0)),$$

上式两边除以 λ, 然后取极限, 注意到

$$\lambda x + (1-\lambda)x_0 = x_0 + \lambda(x - x_0),$$

则左端极限为

$$\nabla f(x_0)^{\mathrm{T}}(x - x_0),$$

式 (2.4.1) 得证.

另一方面, 假设式 (2.4.1) 成立. 对任意 $x_1, x_2 \in S$, $0 < \lambda < 1$, 记 $x_0 = \lambda x_1 + (1-\lambda)x_2$, 由式 (2.4.1) 得

$$f(x_i) \geqslant f(x_0) + \nabla f(x_0)^{\mathrm{T}}(x_i - x_0), \quad i = 1, 2. \tag{2.4.3}$$

对上式进行凸组合得

$$\lambda f(x_1) + (1-\lambda) f(x_2) \geqslant f(x_0) + \nabla f(x_0)^{\mathrm{T}}(\lambda x_1 + (1-\lambda)x_2 - x_0),$$

即

$$f(\lambda x_1 + (1-\lambda)x_2) \leqslant \lambda f(x_1) + (1-\lambda) f(x_2),$$

这说明 $f(x)$ 是凸函数.

(2) 充分性类似于结论 (1) 的证明, 只考虑必要性. 假设对于 $x_0, x_1 \in S$, $x_0 \neq x$, $0 < \lambda < 1$, 式 (2.4.1) 中严格不等式成立, 类似于结论 (1) 的证明可得

$$f(\lambda x_0 + (1-\lambda)x) < \lambda f(x_0) + (1-\lambda) f(x),$$

即 $f(x)$ 是严格凸的.

(3) 根据命题 2.1.1, $f(x)$ 是强凸的充要条件是 $f(x) - \dfrac{1}{2}\|x\|^2$ 是凸的, 对函数 $f(x) - \dfrac{1}{2}\|x\|^2$ 利用结论 (1), 则得结论 (3). 定理得证.

定理 2.4.1 说明, $f(x)$ 是凸函数, 则它的上图总是在它的切平面

$$y = f(x_0) + \nabla f(x_0)^{\mathrm{T}}(x - x_0)$$

之上 (在点 $(x_0, f(x_0))$ 处相交); 当 $f(x)$ 是严格凸时, 只在点 $(x_0, f(x_0))$ 处相交; 如果 $f(x)$ 强凸的, 则它的上图在二次凸函数

$$x \to f(x_0) + \nabla f(x_0)^{\mathrm{T}}(x - x_0) + \frac{1}{2}c\|x - x_0\|^2$$

之上. 这一性质是凸函数最典型的几何特征.

下面讨论可微凸函数梯度的单调性.

定义 2.4.1　设 $S \subset \mathbf{R}^n$, $F(x)$ 为 S 到 $\mathbf{R}^n$ 上的映射, x, x' 为 S 中任意两点, 如果

$$(F(x) - F(x'))^{\mathrm{T}}(x - x') \geqslant 0,$$

则称映射 $F(x)$ 在 S 上是单调的; 如果

$$(F(x) - F(x'))^{\mathrm{T}}(x - x') > 0,$$

则称映射 $F(x)$ 在 S 上是严格单调的; 如果

$$(F(x) - F(x'))^{\mathrm{T}}(x - x') \geqslant c\|x - x_0\|^2,$$

其中 $c>0$, 则称映射 $F(x)$ 在 S 上是强单调 (关于系数 c) 的.

对于一维情形, 单调性退化为单调增加. 下面我们将利用函数梯度的单调性来研究函数自身的凸性.

定理 2.4.2 设 $f(x)$ 为 $\mathbf{R}^n$ 上的可微函数, $S\subset\mathbf{R}^n$ 为凸集, 于是 $f(x)$ 是凸 (严格凸、强凸 (关于参数 c)) 的当且仅当它的梯度 $\nabla f(x)$ 在 S 上是单调 (严格单调、强单调 (关于参数 c)) 的.

证明 必要性. 将凸性与强凸性证明结合在一起, 当参数 $c=0$, 强凸性就是凸性. 假设 $f(x)$ 在集合 S 上是强凸的, 由定理 2.4.1, 对任意 $x_0,x\in S$, 有

$$f(x)\geqslant f(x_0)+\nabla f(x_0)^{\mathrm{T}}(x-x_0)+\frac{1}{2}c\left\|x-x_0\right\|^2,$$
$$f(x_0)\geqslant f(x)+\nabla f(x)^{\mathrm{T}}(x_0-x)+\frac{1}{2}c\left\|x_0-x\right\|^2.$$

上述两式相加, 即可得到 $\nabla f(x)$ 的强单调性.

充分性. 假设 $\nabla f(x)$ 是强单调的. 给定 $x_0,x_1\in\mathbf{R}^n$, 考虑单变量函数 $\varphi(t)=f(x_t)$, 其中 $x_t=x_0+t(x_1-x_0)$, $t\in[0,1]$. 显然, $\varphi(t)$ 是可微的, 其导数为 $\varphi'(t)=\nabla f(x_t)^{\mathrm{T}}(x-x_0)$. 于是, 对于 $0\leqslant t_1<t\leqslant 1$, 有

$$\begin{aligned}\varphi'(t)-\varphi'(t_1)&=(\nabla f(x_t)-\nabla f(x_{t_1}))^{\mathrm{T}}(x_1-x_0)\\&\leqslant\frac{1}{t-t_1}(\nabla f(x_t)-\nabla f(x_{t_1}))^{\mathrm{T}}(x_t-x_{t_1}).\end{aligned}\tag{2.4.4}$$

$\nabla f(x)$ 的单调性和式 (2.4.4) 说明, $\varphi'(t)$ 是单调增加函数, 根据一维函数的有关结果, 函数 $\varphi(t)$ 是凸的.

下面讨论强凸性. 在式 (2.4.4) 中, 令 $t_1=0$, 并考虑到 $\nabla f(x)$ 的强单调性, 得

$$\varphi'(t)-\varphi'(0)\geqslant\frac{1}{t}c\left\|x_t-x_0\right\|^2=tc\left\|x_1-x_0\right\|^2.$$

注意到函数 $\varphi(t)$ 是自身导数的积分, 于是得到

$$\varphi(1)-\varphi(0)-\varphi'(0)=\int_0^1(\varphi'(t)-\varphi'(0))\mathrm{d}t\geqslant c\left\|x_1-x_0\right\|^2,$$

由函数 $\varphi(t)$ 的定义, 上式正是式 (2.4.2), 即 $f(x)$ 是强凸的.

用类似的方法可以证明结论 (2). 定理得证.

第3章　凸函数的次微分

本章讨论非光滑凸函数的微分学. 尽管凸函数在其定义域的内部一定是连续的, 而方向导数也是存在的, 但还不是处处可微的. 在这一章中, 作为可微凸函数梯度概念的推广, 我们引入次微分的概念, 并讨论其有关性质. 本章假定凸函数仅取有限值.

3.1　凸函数次微分的定义及有关性质

定义 3.1.1　设 $f(x)$ 是 $\mathbf{R}^n$ 上的实值函数, 如果极限

$$\lim_{t\to 0^+}\frac{f(x+td)-f(x)}{t} \tag{3.1.1}$$

存在, 则称这个极限是 $f(x)$ 在 x 点沿方向 d 的方向导数, 记为 $f'(x;d)$.

对凸函数, 由于差商 $\dfrac{f(x+td)-f(x)}{t}$ 关于 t 是单调减小的, 另一方面凸函数一定是局部 Lipschitz 的, 记 L 为其 Lipschitz 常数, 则当 t 充分小时, 有

$$\left|\frac{f\left(x+td\right)-f\left(x\right)}{t}\right|\leqslant L\left\|d\right\|,$$

这说明差商 $\dfrac{f\left(x+td\right)-f\left(x\right)}{t}$ 是有界的. 所以极限 (3.1.1) 总是存在的, 也就是说, 凸函数方向导数是存在的, 并且可表示为下述形式:

$$f'(x;d)=\inf_{t>0}\frac{f(x+td)-f(x)}{t}.$$

定理 3.1.1　设 $f(x)$ 为 $\mathbf{R}^n$ 上凸函数, 则对固定的 x, 它的方向导数 $f'(x;d)$ 关于 d 是次线性函数.

证明　设 $d_1,d_2\in\mathbf{R}^n$, $\lambda>0$, 由函数 $f(x)$ 的凸性得

$$\begin{aligned}
&f(x+t(\lambda d_1+(1-\lambda)d_2))-f(x)\\
&=f(\lambda(x+td_1)+(1-\lambda)(x+td_2))-\lambda f(x)-(1-\lambda)f(x)\\
&\leqslant\lambda(f(x+td_1)-f(x))+(1-\lambda)(f(x+td_2)-f(x)),\quad\forall t>0,
\end{aligned}$$

两边同除以 t, 并令 $t\to 0^+$, 则有

$$f'(x;\lambda d_1+(1-\lambda)d_2)\leqslant\lambda f'(x;d_1)+(1-\lambda)f'(x;d_2),$$

这就证明了 $f'(x;d)$ 关于 d 的凸性.

设 $\lambda > 0$, 有

$$\begin{aligned} f'(x;\lambda d) &= \lim_{\tau\to 0^+} \lambda \frac{f(x+\lambda\tau d)-f(x)}{\lambda\tau} \\ &= \lambda \lim_{t\to 0^+} \frac{f(x+td)-f(x)}{t} \\ &= \lambda f'(x;d), \end{aligned}$$

这说明 $f'(x;d)$ 关于 d 是正齐次的, 故 $f'(x;\cdot)$ 是次线性函数. 定理得证.

由于凸函数 $f(x)$ 是局部 Lipschitz 的, 记 L 为 $f(x)$ 在 x 点附近的 Lipschitz 常数, 则有

$$\begin{aligned} |f'(x;d)| &= \left| \lim_{t\to 0^+} \frac{f(x+td)-f(x)}{t} \right| \\ &\leqslant \lim_{t\to 0^+} L\,\|d\| \\ &= L\,\|d\|, \quad \forall d \in \mathbf{R}^n. \end{aligned} \tag{3.1.2}$$

事实上, 不难验证下述不等式成立

$$|f'(x;d_1) - f'(x;d_2)| \leqslant L\,\|d_1 - d_2\|, \quad \forall d_1, d_2 \in \mathbf{R}^n. \tag{3.1.3}$$

定义 3.1.2　设 $f(x)$ 为 $\mathbf{R}^n$ 上凸函数, $f(x)$ 在 x 点的次微分 $\partial f(x)$ 定义如下:

$$\partial f(x) = \left\{ \xi \in \mathbf{R}^n \,\middle|\, f(y) \geqslant f(x) + \xi^{\mathrm{T}}(y-x),\ \forall y \in \mathbf{R}^n \right\},$$

向量 $\xi \in \partial f(x)$ 称为次微分中元素, 也简称为次微分或次梯度.

不等式

$$f(y) \geqslant f(x) + \xi^{\mathrm{T}}(y-x), \quad \forall y \in \mathbf{R}^n \tag{3.1.4}$$

称为次梯度不等式, 它的几何意义是仿射函数 $h(z) = f(x) + \xi^{\mathrm{T}}(z-x)$ 总在 $f(x)$ 上图 Epi$f$ 之下, 且在点 $(x, f(x))$ 处与 Epi$f$ 相交, 因此也称为是 Epi$f$ 在点 $(x, f(x))$ 的一个支撑超平面.

定理 3.1.2　设 $f(x)$ 为 $\mathbf{R}^n$ 上的凸函数, 则 $\xi \in \partial f(x)$ 的充要条件是

$$f'(x;d) \geqslant \xi^{\mathrm{T}} d, \quad \forall d \in \mathbf{R}^n. \tag{3.1.5}$$

证明　必要性. 设 $\xi \in \partial f(x)$, 则由次梯度不等式 (3.1.4) 得

$$\frac{f(x+td)-f(x)}{t} \geqslant \xi^{\mathrm{T}} d, \quad \forall d \in \mathbf{R}^n, t > 0.$$

对上式取极限 ($t \to 0^+$), 即得式 (3.1.5).

充分性. 假设式 (3.1.5) 成立, 因为式 (3.1.1) 中的差商是单减的, 则有

$$\frac{f(x+td)-f(x)}{t} \geqslant f'(x\,;d), \quad \forall d \in \mathbf{R}^n,\, t>0. \tag{3.1.6}$$

考虑到任意 $y \in \mathbf{R}^n$ 都可表示为 $y = x + td$, 将 $y = x + td$ 代入式 (3.1.6) 得式 (3.1.5), 这说明 $\xi \in \partial f(x)$. 定理得证.

定理 3.1.3 设 $f(x)$ 为 $\mathbf{R}^n$ 上的凸函数, 则其次微分 $\partial f(x)$ 为 $\mathbf{R}^n$ 中的凸紧集.

证明 根据定义易见 $\partial f(x)$ 是闭凸集, 为证明紧性, 只需证明 $\partial f(x)$ 的有界性. 假设 $\partial f(x)$ 是无界的, 记 L 为 $f(x)$ 在点 x 附近的 Lipschitz 常数, 则存在 $\xi \in \partial f(x)$, 使得 $\|\xi\| > L$. 根据定理 3.2.1 和式 (3.1.2) 得

$$\left|\xi^{\mathrm{T}} d\right| \leqslant L\,\|\xi\|, \quad \forall d \in \mathbf{R}^n.$$

在上式中选取 $d = \xi$, 则有

$$\|\xi\|^2 = \left|\xi^{\mathrm{T}} d\right| \leqslant L\,\|\xi\|,$$

故 $\|\xi\| \leqslant L$, 这与 $\|\xi\| > L$ 矛盾, 所以 $\partial f(x)$ 是有界的, 进而是凸紧集. 定理得证.

从定理 3.1.2 证明中可以看出下述推论成立.

推论 3.1.1 设 $f(x)$ 为 $\mathbf{R}^n$ 上的凸函数, $L \geqslant 0$ 为 $f(x)$ 在 x 点附近的 Lipschitz 常数, 则对任意 $\xi \in \partial f(x)$, 有 $\|\xi\| \leqslant L$.

下述定理是凸函数微分学中的重要定理之一, 它建立了方向导数和次微分之间的关系.

定理 3.1.4 设 $f(x)$ 为 $\mathbf{R}^n$ 上的凸函数, 则有

$$f'(x\,;d) = \max_{\xi \in \partial f(x)} \xi^{\mathrm{T}} d, \quad \forall d \in \mathbf{R}^n. \tag{3.1.7}$$

证明 根据定理 3.1.3, $\partial f(x)$ 是紧集, 因此式 (3.1.7) 右端 $\max\limits_{\xi \in \partial f(x)} \xi^{\mathrm{T}} d$ 是有意义的. 由定理 3.1.2 可得

$$f'(x\,;d) \geqslant \max_{\xi \in \partial f(x)} \xi^{\mathrm{T}} d, \quad \forall d \in \mathbf{R}^n. \tag{3.1.8}$$

以下证明

$$f'(x\,;d) \leqslant \max_{\xi \in \partial f(x)} \xi^{\mathrm{T}} d, \quad \forall d \in \mathbf{R}^n. \tag{3.1.9}$$

对于给定的 $x, d \in \mathbf{R}^n$, 考虑 $\mathbf{R}^{n+1}$ 中如下子集和半超平面:

$$S_1 = \{(z,w)\,|\,f(z) < w\},$$
$$S_2 = \{(z,w)\,|\,z = x + \alpha d,\, w = f(x) + \alpha f'(x;d),\, \alpha \geqslant 0\}.$$

显然, S_1 和 S_2 是非空凸集, 且不相交, 根据凸集分离定理, 存在非零向量 $(\mu,\gamma)\in\mathbf{R}^{n+1}$, 使得

$$\gamma\left(f\left(x\right)+\alpha f'\left(x;d\right)\right)+\mu^{\mathrm{T}}\left(x+\alpha d\right)\leqslant\gamma w+\mu^{\mathrm{T}}z,\\ \forall\alpha\geqslant 0,\quad z\in\mathbf{R}^n,\quad w>f\left(z\right).\tag{3.1.10}$$

式 (3.1.10) 中 γ 不能小于零, 否则, 式 (3.1.10) 右端当 w 充分大时可任意小, 不等式不能成立, γ 也不可能为零, 否则, 由式 (3.1.10) 可得 $\mu=0$, 与 (μ,γ) 非零相矛盾, 因此必有 $\gamma\geqslant 0$. 不妨假设 $\gamma=1$(当然也可对式 (3.1.10) 两端同除以 γ), 则有

$$f\left(x\right)+\alpha f'\left(x;d\right)+\alpha d^{\mathrm{T}}\mu\leqslant w+\left(z-x\right)^{\mathrm{T}}\mu,\\ \forall\alpha\geqslant 0,\quad z\in\mathbf{R}^n,\quad w>f\left(z\right),\tag{3.1.11}$$

令 $\alpha=0$, 取极限 $w\to f\left(z\right)$ 得

$$f\left(x\right)-\left(z-x\right)^{\mathrm{T}}\mu\leqslant f\left(z\right),\quad\forall z\in\mathbf{R}^n,$$

这意味着 $-\mu\in\partial f\left(x\right)$. 在式 (3.1.11) 中, 令 $z=x,\alpha=1$, 并取极限 $w\to f\left(x\right)$, 则有 $f'\left(x;d\right)\leqslant -d^{\mathrm{T}}\mu$, 这说明式 (3.1.9) 成立. 联立式 (3.1.8) 和式 (3.1.9) 得式 (3.1.7). 定理得证.

定理 3.1.4 说明, 方向导数是次微分的支撑函数, 这样凸函数的次微分与其方向导数相互唯一确定.

定理 3.1.5 设 $f(x)$ 为 $\mathbf{R}^n$ 上凸函数, 则有下述结论:

(1) $\xi\in\partial f(x)$ 的充要条件是 $(\xi\,,-1)$ 是 Epi$f$ 在 $(x\,,f(x))$ 处的法方向, 也就是说 Epi$f$ 在 $(x\,,f(x))$ 处的法锥有下述形式:

$$N_{\mathrm{Epi}f}(x\,,f(x))=\left\{(\lambda\xi\,,-\lambda)\ |\xi\in\partial f(x)\,,\ \lambda\geqslant 0\right\}.\tag{3.1.12}$$

(2) $f(x)$ 的上图 Epi$f$ 在 $(x\,,f(x))$ 处的切锥是方向导数函数 $d\to f'(x\,;d)$ 的上图, 即

$$T_{\mathrm{Epi}f}(x\,,f(x))=\left\{(d,r)\ |r\geqslant f'(x\,;d)\right\}.$$

证明 (1) 根据法锥和上图的定义, $(\xi\,,-1)\in N_{\mathrm{Epi}f}(x,f(x))$ 等价于

$$\xi^{\mathrm{T}}(y-x)+(-1)(r-f(x))\leqslant 0,\quad\forall y\in\mathbf{R}^n,\quad r\geqslant f(y),$$

即

$$r\geqslant f(x)+\xi^{\mathrm{T}}(y-x),\quad\forall y\in\mathbf{R}^n,\quad r\geqslant f(y).$$

显然, 上式与式 (3.1.5) 等价, 也就是与 $\xi \in \partial f(x)$ 等价, 再根据法锥定义知式 (3.1.12) 成立. 结论 (1) 成立.

(2) Epi$f$ 的切锥是式 (3.1.12) 给出锥的极锥, 即由满足下述条件的向量 $(d,r) \in \mathbf{R}^n \times \mathbf{R}$ 形式的集合:

$$\lambda\xi^{\mathrm{T}}d + (-\lambda)r \leqslant 0, \quad \forall \xi \in \partial f(x), \quad \lambda \geqslant 0.$$

不妨考虑 $\lambda > 0$ 情形, 上式两边除以 λ 得

$$r \geqslant \max_{\xi \in \partial f(x)} \xi^{\mathrm{T}}d = f'(x\,;d),$$

结论 (2) 成立. 定理得证.

定理 3.1.6　设 $f(x)$ 为 $\mathbf{R}^n$ 上的实值凸函数, 则 $f(x)$ 在 x 点可微的充要条件是其次微分为单点集, 此时有 $\partial f(x) = \{\nabla f(x)\}$.

证明　必要性. 假设 $f(x)$ 在 x 点可微, 给定 $\xi \in \partial f(x)$, 令

$$x_k = x + \frac{1}{k}(\xi - \nabla f(x)),$$

由 $f(x)$ 在 x 点的可微性得

$$f(x_k) = f(x) + \nabla f(x)^{\mathrm{T}}(x_k - x) + o(||x_k - x||), \tag{3.1.13}$$

由 $\xi \in \partial f(x)$ 及次微分定义得

$$-f(x_k) \leqslant -f(x) - \xi^{\mathrm{T}}(x_k - x). \tag{3.1.14}$$

式 (3.13) 与式 (3.14) 相加得

$$\begin{aligned} 0 &\leqslant (\nabla f(x) - \xi)^{\mathrm{T}}(x_k - x) + o(\|x_k - x\|) \\ &= -\frac{1}{k}||\xi - \nabla f(x)||^2 + o\left(\frac{1}{k}||\xi - \nabla f(x)||\right), \end{aligned}$$

两边乘以 k 得

$$||\xi - \nabla f(x)||^2 \leqslant o(||\xi - \nabla f(x)||).$$

上式说明 $||\xi - \nabla f(x)|| = 0$, 即次微分为单点集, 故 $\partial f(x) = \{\nabla f(x)\}$.

充分性. 假设 $\partial f(x)$ 为单点集, 记 $\partial f(x) = \{\xi\}$, 于是

$$f'(x\,;d) = \max_{\xi \in \partial f(x)} \xi^{\mathrm{T}}d = \xi^{\mathrm{T}}d,$$

而 $f'(x\,;d) = \xi^{\mathrm{T}}d$ 说明 $f(x)$ 是可微的, 其梯度为 $\nabla f(x) = \xi$. 定理得证.

定理 3.1.6 说明, 凸函数次微分确实是可微凸函数梯度概念的推广.

3.2 凸函数的极值条件与中值定理

本节介绍用次微分给出凸函数的极值条件和中值定理. 首先讨论一维函数.

引理 3.2.1 设 $f(x)$ 为一维实值凸函数, 则有

$$\partial f(x)=[-f'(x;-1),f'(x;1)]=\left[f'_-(x),f'_+(x)\right],$$

其中 $f'_-(x)$ 和 $f'_+(x)$ 分别为 $f(x)$ 在 x 处的左右方向导数.

证明 对于一维函数只有两个方向 ±1, 因此由定理 3.1.2, $\xi\in\partial f(x)$ 的充要条件是

$$f'(x;1)\geqslant\xi,\quad f'(x;-1)\geqslant-\xi,$$

所以

$$\partial f(x)=[-f'(x;-1),f'(x;1)].$$

注意到

$$f'_-(x)=-f'(x;-1),f'_+(x)=f'(x;1),$$

结论成立. 引理得证.

定理 3.2.1 设 $f(x)$ 为 $\mathbf{R}^n$ 上的实值凸函数, 则 x^* 为 $f(x)$ 最小点的充分必要条件是 $0\in\partial f(x^*)$.

证明 设 $0\in\partial f(x^*)$, 根据次微分定义, 对任 $x\in\mathbf{R}^n$, 有

$$f(x)-f(x^*)\geqslant 0^{\mathrm{T}}(x-x^*),$$

于是

$$f(x)\geqslant f(x^*),$$

x^* 是 $f(x)$ 的最小值点.

另一方面, 设 x^* 为 $f(x)$ 的最小值点, 则有

$$f(x)\geqslant f(x^*),\quad \forall x\in\mathbf{R}^n,$$

即

$$f(x)\geqslant f(x^*)+0^{\mathrm{T}}(x-x^*).$$

根据次微分定义, $0\in\partial f(x^*)$. 定理得证.

定理 3.2.1 是可微情形最优性条件的推广, 当 $f(x)$ 可微时, $\partial f(x)=\{\nabla f(x)\}$, 此时 $0\in\partial f(x)$ 即为 $0=\nabla f(x)$.

所谓凸函数的中值定理就是利用次微分对任意两点 $x\,,y\in\mathbf{R}^n$ 给出 $f(y)-f(x)$ 的一种表示. 类似于光滑函数的中值定理证明, 我们也要借用这样一个一维函数:

$$\varphi(t)=f(ty+(1-t)x),\quad t\in[0\,,1].\tag{3.2.1}$$

$\varphi(t)$ 是 $f(x)$ 在区间 $[x\,,y]$ 上的限制, 不难验证它是一维凸函数.

引理 3.2.2　设 $f(x)$ 为 $\mathbf{R}^n$ 上的凸函数, 则式 (3.2.1) 定义的凸函数 $\varphi(t)$ 的次微分为

$$\partial\varphi(t)=\left\{\xi^{\mathrm{T}}(y-x)\,|\xi\in\partial f(x_t)\right\},\tag{3.2.2}$$

其中 $x_t=ty+(1-t)x$.

证明　直接计算得

$$\varphi'_+(t)=\lim_{\tau\to0^+}\frac{f(x_t+\tau(y-x))-f(x_t)}{\tau}=f'(x_t;y-x),$$

$$\varphi'_-(t)=\lim_{\tau\to0^-}\frac{f(x_t-\tau(y-x))-f(x_t)}{\tau}=-f'(x_t;-(y-x)).$$

考虑到

$$f'(x_t;y-x)=\max_{\xi\in\partial f(x_t)}\xi^{\mathrm{T}}(y-x),$$

$$-f'(x_t;-(y-x))=-\max_{\xi\in\partial f(x_t)}\xi^{\mathrm{T}}(-y+x)=\min_{\xi\in\partial f(x_t)}\xi^{\mathrm{T}}(y-x),$$

于是

$$\begin{aligned}\partial\varphi(t)&=[\varphi'_-(t)\,,\varphi'_+(t)]\\&=\left[\min_{\xi\in\partial f(x_t)}\xi^{\mathrm{T}}(y-x)\,,\max_{\xi\in\partial f(x_t)}\xi^{\mathrm{T}}(y-x)\right]\\&=\left\{\xi^{\mathrm{T}}(y-x)\,|\xi\in\partial f(x)\right\}.\end{aligned}$$

引理得证.

定理 3.2.2　设 $f(x)$ 为 $\mathbf{R}^n$ 上的实值凸函数, 给定 $x,y\in\mathbf{R}^n$, 且有 $x\neq y$, 则存在 $t\in(0,\ 1)$ 和 $\xi\in\partial f(x_t)$, 其中 $x_t=ty+(1-t)x$, 使得

$$f(y)-f(x)=\xi^{\mathrm{T}}(y-x),\tag{3.2.3}$$

或记为

$$f(y)-f(x)\in\bigcap_{0<t<1}\left\{\xi^{\mathrm{T}}(y-x)\,|\xi\in\partial f(x_t)\right\}.$$

证明　针对式 (3.2.1) 给出的函数 $\varphi(t)$, 引入下述辅助函数:

$$\psi(t)=\varphi(t)-\varphi(0)-t(\varphi(1)-\varphi(0)).$$

显然 $\psi(t)$ 是凸函数, 通过计算 $\psi(t)$ 的左右导数, 得其次微分为

$$\partial\psi(t) = \partial\varphi(t) - (\varphi(1) - \varphi(0)).$$

注意到 $\psi(0) = \psi(1) = 0$, 又 $\psi(t)$ 是连续的, 于是它在某一点 $t \in (0, 1)$ 处达到极小值, 根据定理 3.2.1, $0 \in \partial\psi(t)$, 由引理 3.2.1, 存在 $\xi \in \partial f(x_t)$, 使得

$$\xi^{\mathrm{T}}(y - x) = \varphi(1) - \varphi(0) = f(y) - f(x).$$

定理得证.

中值定理也可以用下述积分形式给出:

$$f(y) - f(x) = \int_0^1 \xi^{\mathrm{T}}(y - x)\mathrm{d}t, \quad \xi \in \partial f(x_t), \tag{3.2.4}$$

式 (3.2.4) 的含义是: 如果 ξ 是 $f(x)$ 在线段 $[x, y]$ 上点 x_t 次微分中的任意一个选择, 式 (3.2.4) 右端的积分与选择无关, 它总是等于 $f(y) - f(x)$. 当 $f(x)$ 是可微凸函数时, 定理 3.2.2 给出的中值定理就是微积分中的 Lagrange 中值定理.

3.3 一些凸函数的次微分

对于一般的凸函数, 其次微分是很难计算的, 特别是解析表达式很难得到, 本节对一些特殊的凸函数, 给出它们次微分的解析表达式.

3.3.1 支撑函数

设 S 为 $\mathbf{R}^n$ 中的非空凸紧集, S 的支撑函数 $\delta_S^*(x)$ 在 $x = 0$ 处的次微分为集合 S 本身, 即 $\partial\delta_S^*(0) = S$. 事实上,

$$\delta_S^*(x) \geqslant \delta_S^*(0) + \xi^{\mathrm{T}}(x - 0), \quad \forall x \in \mathbf{R}^n \tag{3.3.1}$$

等价于

$$\max_{\nu\in S} \nu^{\mathrm{T}}x \geqslant \xi^{\mathrm{T}}x, \quad \forall x \in \mathbf{R}^n. \tag{3.3.2}$$

显然, 如果 $\xi \in S$, 则式 (3.3.2) 成立, 进而式 (3.3.1) 成立, 根据次微分定义 $\xi \in \partial\delta_S^*(0)$, 这说明 $S \subset \partial\delta_S^*(0)$. 另一方面, 设 $\xi \in \partial\delta_S^*(0)$, 则式 (3.3.1) 成立, 进而式 (3.3.2) 成立. 注意到, 式 (3.3.2) 右端可看成为单点集 $\{\xi\}$ 的支撑函数, 根据凸紧集与支撑函数的关系, 由式 (3.3.2) 得 $\xi \in S$, 这说明 $\partial\delta_S^*(0) \subset S$, 于是 $\partial\delta_S^*(0) = S$.

3.3.2 距离函数

假设 S 为 $\mathbf{R}^n$ 中闭凸集, 那么点 x 到 S 的距离函数 $d_S(x) = \min\limits_{y\in S} ||y - x||$ 是凸函数, 它的次微分为

$$\partial d_S(x)=\begin{cases} N_S(x)\bigcap B(0,\ 1), & x\in S,\\ \left\{\dfrac{x-P_S(x)}{||x-P_S(x)||}\right\}, & x\notin S.\end{cases} \tag{3.3.3}$$

为证明公式 (3.3.3), 我们分两种情况:

(1) $x\notin S$, 此时 $d_S(x)>0$, 利用链锁法则有

$$\nabla d_S(x)=\nabla\sqrt{d_S^2(x)}=\frac{\nabla d_s^2(x)}{2d_S(x)}.$$

于是, 只要证明

$$\nabla d_s^2(x)=2(x-p_S(x)),$$

其中 $p_S(x)$ 为点 x 到集合 S 的投影. 记

$$\Delta=d_s^2(x+h)-d_s^2(x),$$

由于

$$d_s^2(x)\leqslant ||x-p_S(x+h)||^2,$$

则有

$$\begin{aligned}\Delta&\geqslant ||x+h-p_S(x+h)||^2-||x-p_S(x+h)||^2\\ &=||h||^2+2h^{\mathrm{T}}(x-p_S(x+h)).\end{aligned} \tag{3.3.4}$$

x 与 $x+h$ 互换位置, 可得

$$\Delta\leqslant ||x+h-p_S(x)||^2-||x-p_S(x)||^2=||h||^2+2h^{\mathrm{T}}(x-p_S(x)). \tag{3.3.5}$$

因为函数 $p_S(x)$ 是非膨胀的 (简式 (1.2.4)), 即

$$\|p_S(x_1)-p_S(x_2)\|\leqslant\|x_1-x_2\|,$$

于是

$$p_S(x+h)=p_S(x)+o(\|h\|),$$

将此式代入式 (3.3.4) 后联立式 (3.3.4) 和式 (3.3.5) 得

$$\Delta=2h^{\mathrm{T}}(x-p_S(x))+o(\|h\|),$$

于是

$$\nabla d_s^2(x)=2(x-p_S(x)).$$

(2) 设 $x \in S$, 记 $\xi \in \partial d_S(x)$, 即

$$d_S(x') \geqslant \xi^{\mathrm{T}}(x' - x), \quad \forall x' \in \mathbf{R}^n,$$

这意味着

$$\xi^{\mathrm{T}}(x' - x) \leqslant 0, \quad \forall x' \in S,$$

于是 $\xi \in N_S(x)$. 选取 $x' = x + \xi$, 有

$$||\xi||^2 \leqslant d_S(x + \xi) \leqslant ||x + \xi - x|| = ||\xi||.$$

反过来, 设 $\xi \in N_S(x) \bigcap B(0, 1)$, 考虑

$$\xi^{\mathrm{T}}(x' - x) = \xi^{\mathrm{T}}(x' - p_S(x')) + \xi^{\mathrm{T}}(p_S(x') - x),$$

由于 $\xi \in N_S(x)$, 上式中最后一项是非膨胀的, 由于 Cauchy-Schwarz 不等式及 $||\xi|| \leqslant 1$, 得

$$\xi^{\mathrm{T}}(x' - p_S(x')) \leqslant ||x' - p_S(x')|| = d_S(x'),$$

故有 $\xi \in \partial d_S(x)$. 定理得证.

3.3.3 复合函数

定理 3.3.1 设 $f_1(x)$ 和 $f_2(x)$ 为 $\mathbf{R}^n$ 上的凸函数, 则对任意 $\lambda_1, \lambda_2 \geqslant 0$, 有

$$\partial(\lambda_1 f_1(x) + \lambda_2 f_2(x)) = \lambda_1 \partial f_1(x) + \lambda_2 \partial f_2(x).$$

证明 直接计算 $\lambda_1 f_1(x) + \lambda_2 f_2(x)$ 的方向导数, 得

$$\begin{aligned}(\lambda_1 f_1(x) + \lambda_2 f_2(x))'(x\,;d) &= \lambda_1 f_1'(x\,;d) + \lambda_2 f_2'(x\,;d) \\ &= \lambda_1 \max_{\xi \in \partial f_1(x)} \xi^{\mathrm{T}} d + \lambda_2 \max_{\xi \in \partial f_2(x)} \xi^{\mathrm{T}} d \\ &= \max\left\{\xi^{\mathrm{T}} d \,|\, \xi \in \lambda_1 \partial f_1(x) + \lambda_2 \partial f_2(x)\right\}, \quad (3.3.6)\end{aligned}$$

上式中最后两步利用了支撑函数的性质. 式 (3.3.6) 说明, 函数 $\lambda_1 f_1(x) + \lambda_2 f_2(x)$ 方向导数的支撑函数为 $\lambda_1 \partial f_1(x) + \lambda_2 \partial f_2(x)$, 于是它的次微分为 $\lambda_1 \partial f_1(x) + \lambda_2 \partial f_2(x)$. 定理得证.

定理 3.3.1 是关于次微分链锁法则, 但由于凸性只对加法和正数乘封闭, 因此, 对于次微分, 没有乘、除以及复合运算等一般形式的运算法则.

定理 3.3.2 设 A 为 $\mathbf{R}^n$ 到 $\mathbf{R}^m$ 上的仿射映射, 记 $A(x) = A_0 x + b$, 其中 A_0 为 $m \times n$ 阶矩阵, $b \in \mathbf{R}^m$, $g(y)$ 是 $\mathbf{R}^m$ 上的凸函数, 则复合函数 $g \circ A(x)$ 的次微分为

$$\partial(g \circ A)(x) = A_0^{\mathrm{T}} \partial g(A(x)), \quad \forall x \in \mathbf{R}^n. \quad (3.3.7)$$

证明 计算 $(g \circ A)(x)$ 的方向导数

$$
\begin{aligned}
(g \circ A)'(x\,;d) &= \lim_{t \to 0^+} \frac{(g \circ A)(x+td) - (g \circ A)(x)}{t} \\
&= \lim_{t \to 0^+} \frac{g(A(x)+tA_0 d) - g(A(x))}{t} \\
&= g'(A(x)\,;A_0 d), \quad \forall d \in \mathbf{R}^n.
\end{aligned} \tag{3.3.8}
$$

上式右端可表示为

$$
\begin{aligned}
g'(A(x)\,;A_0 d) &= \max\{\xi^{\mathrm{T}} A_0 d | \xi \in \partial g(A(x))\} \\
&= \max\{\xi^{\mathrm{T}} d | \xi \in A_0^{\mathrm{T}} \partial g(A(x))\}.
\end{aligned} \tag{3.3.9}
$$

联立式 (3.3.8) 和式 (3.3.9) 说明, $(g \circ A)(x)$ 的方向导数为 $A_0^{\mathrm{T}} \partial A(x)$ 的支撑函数, 于是它的次微分为 $A_0^{\mathrm{T}} \partial g(A(x))$. 定理得证.

3.3.4 极大值函数

定义 3.3.1 设 $f(x)$ 为 $\mathbf{R}^n$ 上的函数, $x_0 \in \mathbf{R}^n$, 如果对任意 $\varepsilon > 0$, 存在常数 δ, 使得

$$
f(x) - f(x_0) < \varepsilon, \quad \forall x \in B(x, \delta),
$$

则称 $f(x)$ 在 x_0 处是上半连续的; 如果对任意 $\varepsilon > 0$, 存在常数 δ, 使得

$$
-\varepsilon < f(x) - f(x_0), \quad \forall x \in B(x, \delta),
$$

则称 $f(x)$ 在 x_0 处是下半连续的.

$f(x)$ 在 x_0 处连续当且仅当 $f(x)$ 在 x_0 处同时上半连续和下半连续.

考虑函数

$$
f(x) = \sup\{f_i(x) | i \in I\}, \tag{3.3.10}
$$

其中 I 是任意一指标集, $f_i(x)$ 为 $\mathbf{R}^n$ 上的凸函数, 再假设 $f(x) < +\infty$. 我们知道, 式 (3.3.10) 给出的函数 $f(x)$ 是凸函数.

命题 3.3.1 对于式 (3.3.10) 给出的函数 $f(x)$, 其次微分有关系式:

$$
\operatorname{cl}\operatorname{co}\{\bigcup \partial f_i(x) | i \in I(x)\} \subset \partial f(x), \tag{3.3.11}
$$

其中

$$
I(x) = \{i \in I | f_i(x) = f(x)\}.
$$

证明 给定 $i \in I(x)$ 和 $\xi \in \partial f_i(x)$, 根据次微分定义有

$$
\begin{aligned}
f(y) \geqslant f_i(y) &\geqslant f_i(x) + \xi^{\mathrm{T}}(y - x) \\
&= f(x) + \xi^{\mathrm{T}}(y - x), \quad \forall y \in \mathbf{R}^n.
\end{aligned}
$$

上式说明 $\xi \in \partial f(x)$, 故

$$\partial f_i(x) \subset \partial f(x), \quad i \in I(x).$$

再由 $\partial f(x)$ 的闭凸性, $\partial f(x)$ 包含 $\partial f_i(x), i \in I(x)$ 的并及其闭凸包, 式 (3.3.9) 成立. 命题得证.

式 (3.3.11) 仅仅是一种包含关系, 等式不一定成立, 下述定理在一定条件下得到了等式关系.

定理 3.3.3 假设指标集 I 为紧集 (在某一度量空间里), 且对每一固定 $x \in \mathbf{R}^n$, 函数 $i \to f_i(x)$ 是上半连续的, 则有

$$\partial f(x) = \operatorname{co}\{\bigcup \partial f_i(x) | i \in I(x)\}. \tag{3.3.12}$$

证明 已知条件 $f(x) < +\infty$ 和 I 的紧性及 $i \to f_i(x)$ 的上半连续性说明, $I(x)$ 是非空紧集. 记

$$S = \bigcap_{i \in I(x)} \partial f_i(x),$$

根据式 (3.3.11), S 是有界的, 以下讨论 S 的闭性. 设 $\{s_k\} \subset S$ 为一收敛点列, 其极限为 s, 于是对每一个 s_k 存在指标 $i_k \in I(x)$, 使得 $s_k \in \partial f_{i_k}(x)$, 即

$$f_{i_k}(y) \geqslant f_{i_k}(x) + s_k(y - x), \quad \forall y \in \mathbf{R}^n.$$

选取 $\{i_k\}$ 中一收敛子列, 不妨记为 $\{i_k\}$ 本身, 令 $k \to +\infty$, 根据 $I(x)$ 的紧性得 $i_k \to i \in I(x)$, 这时有

$$f_{i_k}(x) \equiv f(x) = f_i(x).$$

再由函数 $f_{(\cdot)}(y)$ 的上半连续性, 得

$$f_i(y) \geqslant \lim_{k \to \infty} \sup f_{i_k}(y) \geqslant f_i(x) + s^{\mathrm{T}}(y - x), \quad \forall y \in \mathbf{R}^n,$$

于是 $s \in \partial f_i(x) \subset S$, 这就证明了 S 是闭的. 另外, $s \in \partial f_i(x) \subset S$ 和 S 的紧性说明它的凸包也是紧的, 于是式 (3.3.12) 的右端也是紧集. 根据命题 3.3.1, 只需证明

$$\partial f(x) \subset \operatorname{co}\{\bigcup \partial f_i(x) | i \in I(x)\}. \tag{3.3.13}$$

为此要证明下述不等式 (其中等式部分由支撑函数得到):

$$f'(x\,; d) \leqslant \delta_S(d) = \sup\{f_i'(x\,; d) | i \in I(x)\}. \tag{3.3.14}$$

对于 $\varepsilon > 0$, 有

$$\frac{f(x + td) - f(x)}{t} > f'(x\,; d) - \varepsilon, \quad \forall t > 0. \tag{3.3.15}$$

对于 $t>0$, 令

$$I_t=\left\{i\in I\left|\frac{f_i(x+td)-f(x)}{t}\geqslant f'(x\,;d)-\varepsilon\right.\right\},$$

显然 I_t 是非空集. 再由 I 是紧集以及函数 $f_{(\cdot)}(x+td)$ 是上半连续的, 可知 I_t 为紧集. 事实上, I_t 是函数

$$t\to\frac{f_i\,(x+td)-f_i\,(x)}{t}+\frac{f_i\,(x)-f\,(x)}{t}$$

的水平集, 其中第一项为凸函数的差商, 第二项分子非正, 因此对于 $0<t_1\leqslant t_2$, 有 $I_{t_1}\subset I_{t_2}$.

由 I_t 的紧性, 必存在 $i^*\in\bigcap\limits_{t>0}I_t$(对每一个 $\tau\in(0,t)$), 选取 $i_\tau\in I_\tau\subset I_t,\tau\to 0^+$ 的聚点, 它就在 I_t 中, 于是有

$$f_{i^*}(x+td)-f(x)\geqslant t(f'(x\,;d)-\varepsilon),\quad\forall t>0,$$

因此 $i^*\in I(x)$(凸函数 f_{i^*} 的连续性, $t\to 0^+$). 在上述不等式中, 用 $f_{i^*}(x)$ 代替 $f(x)$, 并除以 t, 而后再取极限得

$$\delta_S(d)\geqslant f'_{i^*}(x\,;d)\geqslant f'(x\,;d)-\varepsilon.$$

由 $d\in\mathbf{R}^n$ 和 $\varepsilon>0$ 的任意性, 式 (3.3.14) 成立. 定理得证.

推论 3.3.1　设 $Y\subset\mathbf{R}^p$ 为紧集, $g\,(x)$ 为定义于 $\mathbf{R}^n\times Y$ 上的函数且满足下述条件:

(1) 对每个 $x\in\mathbf{R}^n$, $g\,(x,\cdot)$ 是上半连续的;

(2) 对每个 $y\in Y$, $g\,(\cdot,y)$ 是可微凸函数;

(3) 函数 $f\,(x)=\max\limits_{y\in Y}f\,(x,y)$ 存在.

记

$$Y\,(x)=\left\{y=f\,(x,y)=\max_{y\in Y}f\,(x,y)\right\},$$

则有

$$\partial f\,(x)=\mathrm{co}\,\{\nabla_x f\,(x,y)\,|y\in Y\,(x)\}\,.$$

推论 3.3.2　设 $f_i(x)$, $i\in I$ 为 $\mathbf{R}^n$ 上的可微凸函数, I 为有限指标集, 则极大值函数 $f(x)=\max\limits_{i\in I}f_i(x)$ 的次微分为

$$\partial f(x)=\mathrm{co}\{\nabla f_i(x)|i\in I(x)\},\tag{3.3.16}$$

其中

$$I(x)=\{i\in I|f_i(x)=f(x)\}.$$

证明 对固定的 x, 函数 $i \to f_i(x)$ 变量只取离散值, 因此一定是上半连续的, 由定理 3.3.3, 式 (3.3.16) 成立. 推论得证.

推论 3.3.2 中的有限可微凸函数的极大值函数是非光滑优化中一种最为常见的凸函数.

3.4 次微分的单调性和连续性

3.4.1 单调性

第 2 章已经讨论过可微凸函数梯度的单调性, 本节将其推广到非光滑凸函数的次微分形式.

定理 3.4.1 设 $f(x)$ 为 $\mathbf{R}^n$ 上的凸函数, 则次微分映射 $x \to \partial f(x)$ 是单调的, 即对任意 $x_1,\ x_2 \in \mathbf{R}^n$, 有

$$(\xi_2 - \xi_1)^{\mathrm{T}}(x_2 - x_1) \geqslant 0, \quad \forall \xi_i \in \partial f(x_i), \quad i = 1,\ 2. \tag{3.4.1}$$

证明 考虑次梯度不等式:

$$f(x_2) \geqslant f(x_1) + \xi_1^{\mathrm{T}}(x_2 - x_1), \quad \forall \xi_1 \in \partial f(x_1)$$

$$f(x_1) \geqslant f(x_2) + \xi_2^{\mathrm{T}}(x_2 - x_1), \quad \forall \xi_2 \in \partial f(x_2),$$

将以上两式相加即得到式 (3.4.1). 定理得证.

定理 3.4.2 设 $f(x)$ 为 $\mathbf{R}^n$ 上的函数, 则 $f(x)$ 是强凸 (关于常数 c) 的充要条件是对任意 $x_1,\ x_2 \in \mathbf{R}^n$, 有

$$f(x_2) \geqslant f(x_1) + \xi^{\mathrm{T}}(x_2 - x_1) + \frac{1}{2}c||x_2 - x_1||^2, \quad \forall \xi \in \partial f(x_1), \tag{3.4.2}$$

或等价地

$$(\xi_2 - \xi_1)^{\mathrm{T}}(x_2 - x_1) \geqslant c||x_2 - x_1||^2, \quad \forall \xi_i \in \partial f(x_i), \quad i = 1, 2. \tag{3.4.3}$$

证明 充分性. 给定 $x_1,\ x_2 \in \mathbf{R}^n$, $\lambda \in (0,\ 1)$, 记

$$x_\lambda = \lambda x_2 + (1 - \lambda)x_1 = x_1 + \lambda(x_2 - x_1).$$

假设式 (3.4.2) 成立, 则对任意 $\xi \in \partial f(x_\lambda)$, 有

$$f(x_2) \geqslant f(x_\lambda) + (1 - \lambda)\xi^{\mathrm{T}}(x_2 - x_1) + \frac{1}{2}c(1 - \lambda)^2||x_2 - x_1||^2,$$

$$f(x_1) \geqslant f(x_\lambda) + \lambda\xi^{\mathrm{T}}(x_1 - x_2) + \frac{1}{2}c\lambda^2||x_1 - x_2||^2,$$

以上两式分别乘以 λ 和 $(1-\lambda)$, 然后相加得

$$\lambda f(x_2)+(1-\lambda)f(x_1)\geqslant f(x_\lambda)+\frac{1}{2}c\|x_2-x_1\|^2(\lambda(1-\lambda)^2+(1-\lambda)\lambda^2). \tag{3.4.4}$$

注意到 $\lambda(1-\lambda)^2+(1-\lambda)\lambda^2$ 有界且不为零, 式 (3.4.4) 说明 $f(x)$ 是强凸的.

必要性. 假设 $f(x)$ 是强凸的, 即对任意 $x_1,\ x_2\in\mathbf{R}^n$, $\lambda\in(0,\ 1)$, 有

$$\begin{aligned}f(x_\lambda)&=f(\lambda x_2+(1-\lambda)x_1)\\&\leqslant\lambda f(x_2)+(1-\lambda)f(x_1)-\frac{1}{2}c\|x_2-x_1\|\\&\leqslant\lambda f(x_2)+(1-\lambda)f(x_1)-\frac{1}{2}c\lambda(1-\lambda)\|x_2-x_1\|^2,\end{aligned}$$

上式变形为

$$\frac{f(x_\lambda)-f(x_1)}{\lambda}+\frac{1}{2}c(1-\lambda)\|x_2-x_1\|^2\leqslant f(x_2)-f(x_1).$$

令 $\lambda\to 0^+$, 得

$$f'(x_1\,;x_2-x_1)+\frac{1}{2}c\|x_2-x_1\|^2\leqslant f(x_2)-f(x_1),$$

进一步可得式 (3.4.2). 定理得证.

关于式 (3.4.2) 与式 (3.4.3) 的等价性, 这里从略.

类似定理 3.4.2 可以得到下述定理.

定理 3.4.3　$\mathbf{R}^n$ 中凸函数 $f(x)$ 是严格凸的充要条件是对任意 $x_1,\ x_2\in\mathbf{R}^n$ 且 $x_1\neq x_2$, 有

$$f(x_2)>f(x_1)+\xi^{\mathrm{T}}(x_2-x_1),\quad \forall\xi\in\partial f(x_1),$$

或等价地

$$(\xi_2-\xi_1)^{\mathrm{T}}(x_2-x_1)>0,\quad \forall\xi_i\in\partial f(x_i),\quad i=1,\ 2.$$

证明类似于定理 3.4.2, 这里从略.

3.4.2　次微分的上半连续性

定义 3.4.1　设 $F(x)$ 为 $\mathbf{R}^n$ 到 $\mathbf{R}^n$ 子集上的集值影射, $x_0\in\mathbf{R}^n$, 如果对任意 $\varepsilon>0$, 存在常数 δ, 使得

$$F(x)\subset F(x_0)+B(x,\varepsilon),\quad \forall x\in B(x,\delta),$$

则称集值影射 $F(x)$ 在 x_0 处是上半连续的; 如果对任意 $\varepsilon>0$, 存在常数 δ, 使得

$$F(x_0)\subset F(x)+B(x,\varepsilon),\quad \forall x\in B(x,\delta),$$

则称集值影射 $F(x)$ 在 x_0 处是下半连续的. 同时满足上、下半连续的集值影射称为是 Hausdorff 连续的.

集值影射上、下半连续是非光滑分析中的一个重要概念, 在非光滑优化中通常考虑上半连续, 这是因为许多次微分都是上半连续的; 在非光滑控制理论中上、下半连续都有应用, 甚至有同时满足上、下半连续的集值影射.

例 3.4.1 定义 $\mathbf{R}$ 上的集值影射 $F(x)$ 如下:

$$F(x)=\begin{cases}\{0\}, & x\neq 0,\\ [-1,1], & x=0,\end{cases}$$

不难验证 $F(x)$ 在 $x=0$ 处是上半连续的.

例 3.4.2 定义 $\mathbf{R}$ 上的集值影射 $F(x)$ 如下:

$$F(x)=\begin{cases}[-1,1], & x\neq 0,\\ \{0\}, & x=0.\end{cases}$$

不难验证 $F(x)$ 在 $x=0$ 处是下半连续的.

定理 3.4.4 设 $f(x)$ 为 $\mathbf{R}^n$ 上的凸函数, 则集值映射 $x\to\partial f(x)$ 是上半连续的, 即给定 $x\in\mathbf{R}^n$, 对任意 $\varepsilon>0$, 存在 $\delta>0$, 使得

$$\partial f(x')\subset\partial f(x)+B(0,\varepsilon),\quad \forall x'\in B(x,\delta). \tag{3.4.5}$$

证明 用反证法. 假设式 (3.4.5) 在 x 点处不成立, 则存在

$$\varepsilon>0,\quad x_k\in\mathbf{R}^n,\quad \xi_k\in\partial f(x_k),\quad k=1,2,\cdots,$$

使得 $x_k\to x\ (k\to\infty)$,

$$\xi_k\notin\partial f(x)+B(0,\varepsilon),\quad k=1,2,\cdots. \tag{3.4.6}$$

根据次微分定义有

$$f(y)\geqslant f(x_k)+\xi_k^{\mathrm{T}}(y-x_k). \tag{3.4.7}$$

由次微分的局部有界性, $\{\xi_k\}$ 存在收敛子列, 不妨记为 $\{\xi_k\}$ 本身, 记 $\xi_k\to\xi$, 在式 (3.4.7) 中, 关于 $k\to\infty$ 取极限得

$$f(y)\geqslant f(x)+\xi^{\mathrm{T}}(y-x),$$

这说明 $\xi\in\partial f(x)$, 这与式 (3.4.6) 矛盾. 定理得证.

3.5　ε 次微分和 ε 方向导数

本节引入近似次微分和近似方向导数, 也称为 ε 次微分和 ε 方向导数, 并介绍一些基本性质, 其中部分证明略去, 有兴趣的读者可参阅 Hiriart-Urruty 和 Lemarechal 或 Rockfellar 关于凸分析的专著.

定义 3.5.1　设 $f(x)$ 为 $\mathbf{R}^n$ 上凸函数, ε 为非负常数, $f(x)$ 在 x 点的 ε 次微分 $\partial_\varepsilon f(x)$ 定义如下:

$$\partial_\varepsilon f(x) = \left\{\xi \in \mathbf{R}^n \,\middle|\, f(y) \geqslant f(x) + \xi^{\mathrm{T}}(y-x) - \varepsilon,\ \forall y \in \mathbf{R}^n\right\}, \tag{3.5.1}$$

向量 $\xi \in \partial_\varepsilon f(x)$ 称为 ε 次微分中元素, 也简称为 ε 次微分或 ε 次梯度.

从 ε 次微分定义易见下述关系式成立:

$$\partial_{\varepsilon_1} f(x) \subset \partial_{\varepsilon_2} f(x), \quad \varepsilon_1 \leqslant \varepsilon_2, \tag{3.5.2}$$

$$\partial f(x) = \partial_0 f(x) = \bigcap\{\partial_\varepsilon f(x) \,|\, \varepsilon > 0\}, \tag{3.5.3}$$

$$\partial_{\lambda\varepsilon_1+(1-\lambda)\varepsilon_2} f(x) \supset \lambda\partial_{\varepsilon_1} f(x) + (1-\lambda)\partial_{\varepsilon_2} f(x), \quad 0 < \lambda < 1. \tag{3.5.4}$$

式 (3.5.3) 说明次微分是 ε 次微分的一个特殊情况.

定理 3.5.1　设 $f(x)$ 为 $\mathbf{R}^n$ 上的有限凸函数, $\varepsilon > 0$, 则 ε 次微分 $\partial_\varepsilon f(x)$ 为非空集.

证明　根据凸集的分离定理, 存在一个超平面严格分离点 $(x, f(x)-\varepsilon)$ 和凸集 Epi$f$, 于是存在常数 α 和 $s \in \mathbf{R}^n$, 使得

$$s^{\mathrm{T}}x + \alpha(f(x)-\varepsilon) < s^{\mathrm{T}}y + \alpha f(x), \quad y \in \mathbf{R}^n. \tag{3.5.5}$$

在式 (3.5.5) 中取 $y = x$, 则有

$$\alpha(f(x)-\varepsilon) < \alpha f(x) < +\infty,$$

这就说明 $\alpha > 0$, 即超平面不是垂直的, 这样

$$s' = \frac{s}{\alpha} \in \partial_\varepsilon f(x).$$

定理得证.

定理 3.5.2　设 $f(x)$ 为 $\mathbf{R}^n$ 上的有限凸函数, $\varepsilon \geqslant 0$, 则 ε 次微分 $\partial_\varepsilon f(x)$ 是 $\mathbf{R}^n$ 中的凸紧集.

证明 闭性和凸性可以从定义中直接得到, 我们只需证明有界性. $f(x)$ 是凸函数, 因此是局部 Lipschitz 的, 设 L 是 $f(x)$ 在 $B(x,\delta)$ 内的 Lipschitz 常数, 其中 $\delta>0$. 对于 $\xi\in\partial_\varepsilon f(x)$ 且 $\xi\neq 0$, 选取

$$y=x+\frac{\delta}{\|\xi\|}\xi,$$

根据 ε 次微分的定义, 则有

$$f(x)+L\delta\geqslant f(y)\geqslant f(x)+\frac{\delta}{\|\xi\|}\xi^{\mathrm{T}}\xi-\varepsilon,$$

即

$$\|\xi\|\leqslant L+\frac{\varepsilon}{\delta},$$

这就说明 $\partial_\varepsilon f(x)$ 是有界的. 定理得证.

利用 ε 次微分也可建立相应 ε 最优解的最优性条件. 事实上, 利用 ε 次微分定义可以直接得到

$$0\in\partial_\varepsilon f(x)\ \text{当且仅当}\ f(x)\leqslant f(y)+\varepsilon,\quad \forall y\in\mathbf{R}^n.$$

根据定义可以得到 ε 次微分的一些运算性质.

命题 3.5.1 设 $f(x)$ 为 $\mathbf{R}^n$ 上的有限凸函数, 下述结论成立:

(1) 设 $g(x)=f(x)+c$, 其中 c 为常数, 则有 $\partial_\varepsilon g(x)=\partial_\varepsilon f(x)$;

(2) 设 $g(x)=af(x)$, 其中 $a>0$, 则 $\partial_\varepsilon g(x)=a\partial_{\frac{\varepsilon}{a}}f(x)$;

(6) 设 $g(x)=f(ax)$, 其中 $a\neq 0$, 则 $\partial_\varepsilon g(x)=a\partial_\varepsilon f(ax)$;

(4) 设 $g(x)=f(x-x_0)$, 其中 $x_0\in\mathbf{R}^n$, 则 $\partial_\varepsilon g(x+x_0)=\partial_\varepsilon f(x)$;

(5) 设 $g(x)=f(x)+\alpha^{\mathrm{T}}x$, 其中 $\alpha\in\mathbf{R}^n$, 则 $\partial_\varepsilon g(x)=\partial_\varepsilon f(x)+\{\alpha\}$;

(6) 如果 $f_1(x)\leqslant f_2(x)$, $\forall x\in\mathbf{R}^n$, $f_1(x_0)=f_2(x_0)$, 其中 $x_0\in\mathbf{R}^n$, 则有 $\partial_\varepsilon f_1(x_0)\subset\partial_\varepsilon f_2(x_0)$.

ε 次微分的支撑函数, 称为 ε 方向导数, 记为 $f'_\varepsilon(x;d)$, 它表示为

$$f'_\varepsilon(x;d)=\max_{\xi\in\partial_\varepsilon f(x)}\xi^{\mathrm{T}}d,\quad d\in\mathbf{R}^n. \tag{3.5.6}$$

可以证明 ε 方向导数可以表为下述形式:

$$f'_\varepsilon(x;d)=\inf_{t>0}\frac{f(x+td)-f(x)+\varepsilon}{t}, \tag{3.5.7}$$

利用有关专门知识结合 ε 次微分定义可以得到下述结果.

定理 3.5.3 设 $f_1(x),f_2(x)$ 为 $\mathbf{R}^n$ 上的有限凸函数, $\varepsilon\geqslant 0$, 则

$$\partial_\varepsilon(f_1(x)+f_2(x))\supset\bigcup\{\partial_{\varepsilon_1}f_1(x)+\partial_{\varepsilon_2}f_2(x)\,|\,\varepsilon_1,\varepsilon_2\geqslant 0,\ \varepsilon_1+\varepsilon_2\leqslant\varepsilon\}.$$

第 4 章　局部 Lipschitz 函数的广义梯度

针对局部 Lipschitz 函数建立的广义梯度是对可微函数经典梯度的推广, 它也是目前为止关于非凸函数各种次微分中最有影响的一种, 在非光滑分析中占有重要位置, 在最优化中得到广泛应用. 本章将介绍广义梯度及有关性质.

4.1　广义梯度定义和基本性质

4.1.1　局部 Lipschitz 函数

广义梯度是针对局部 Lipschitz 函数引入的, 因此, 首先介绍局部 Lipschitz 函数.

设 $f_1(x), f_2(x)$ 是 $S \subset \mathbf{R}^n$ 上的局部 Lipschitz 函数, 不难验证对任意实数 λ_1, λ_2, $\lambda_1 f_1(x) + \lambda_2 f_2(x)$ 是 S 上的局部 Lipschitz 函数. 这一事实说明局部 Lipschitz 函数全体构成一个线性空间.

命题 4.1.1　设 $f_1(x), f_2(x)$ 是 $S \subset \mathbf{R}^n$ 上的局部 Lipschitz 函数, 则函数

$$g(x) = \max\{f_1(x), f_2(x)\},$$

$$h(x) = \min\{f_1(x), f_2(x)\}$$

也是 S 上的局部 Lipschitz 函数.

证明　显然, $g(x)$ 和 $h(x)$ 可以表示为下述形式

$$g(x) = \frac{1}{2}(f_1(x) + f_2(x) + |f_1(x) - f_2(x)|),$$

$$h(x) = \frac{1}{2}(f_1(x) + f_2(x) - |f_1(x) - f_2(x)|).$$

设 L_1, L_2 分别为 $f_1(x)$ 和 $f_2(x)$ 在 x 点的 Lipschitz 常数, 记 $L = \max\{L_1, L_2\}$, 直接推导得

$$\begin{aligned}|g(x_1) - g(x_2)| =& \frac{1}{2}|f_1(x_1) + f_2(x_1) + |f_1(x_1) - f_2(x_1)| \\ & -(f_1(x_2) + f_2(x_2) + |f_1(x_2) - f_2(x_2)|)| \\ =& \frac{1}{2}|f_1(x_1) - f_1(x_2) + f_2(x_1) - f_2(x_2) \\ & +|f_1(x_1) - f_2(x_1)| - |f_1(x_2) - f_2(x_2)||\end{aligned}$$

$$
\begin{aligned}
&\leqslant \frac{1}{2}(|f_1(x_1) - f_1(x_2)| + |f_2(x_1) - f_2(x_2)| + |f_1(x_1) - f_1(x_2) + f_2(x_2) - f_2(x_1)|) \\
&\leqslant \frac{1}{2}(|f_1(x_1) - f_1(x_2)| + |f_2(x_1) - f_2(x_2)| + |f_1(x_1) - f_1(x_2)| + |f_2(x_1) - f_2(x_2)|) \\
&= |f_1(x_1) - f_1(x_2)| + |f_2(x_1) - f_2(x_2)| \\
&\leqslant L_1 |x_1 - x_2| + L_2 |x_1 - x_2| \\
&= 2L |x_1 - x_2|,
\end{aligned}
$$

这说明 $g(x)$ 是局部 Lipschitz 函数.

同理可证

$$
\begin{aligned}
|h(x_1) - h(x_2)| &\leqslant |f_1(x_1) - f_1(x_2)| + |f_2(x_1) - f_2(x_2)| \\
&\leqslant 2L |x_1 - x_2|,
\end{aligned}
$$

所以 $h(x)$ 也为局部 Lipschitz 函数. 命题得证.

命题 4.1.1 表明, 局部 Lipschitz 函数关于极大值和极小值运算是封闭的, 这也说明了局部 Lipschitz 函数的广泛性. 常见的优化问题所涉及的函数也多为局部 Lipschitz 的.

4.1.2 广义梯度定义与性质

针对局部 Lipschitz 函数, 在吸收前人一系列研究成果基础上, 作为连续可微函数梯度的推广, Clarke 提出了广义梯度概念.

定义 4.1.1 设 $f(x)$ 为开集 $S \subset \mathbf{R}^n$ 上的局部 Lipschitz 函数, $f(x)$ 在点 x 处关于方向 $d \in \mathbf{R}^n$ 的广义方向导数定义为

$$
f^{\circ}(x; d) = \limsup_{\substack{y \to x \\ t \to 0^+}} \frac{f(y + td) - f(y)}{t}. \tag{4.1.1}
$$

由 $f(x)$ 的局部 Lipschitz 性质, 右端差商总是有界的, 因此上极限一定存在, 所以广义方向导数对于局部 Lipschitz 函数总是存在的. 一般说来, 局部 Lipschitz 函数的经典方向导数

$$
f'(x; d) = \lim_{t \to 0^+} \frac{f(x + td) - f(x)}{t}, \quad d \in \mathbf{R}^n
$$

不一定存在, 即使存在, 一般来讲与广义方向导数也不相等, 但如果存在, 则要比广义方向导数小, 即

$$
f'(x; d) \leqslant f^{\circ}(x; d), \quad \forall d \in \mathbf{R}^n
$$

总是成立的.

例 4.1.1　设

$$f(x)=\begin{cases} x^2\sin\dfrac{1}{x}, & x\neq 0,\\ 0, & x=0,\end{cases}$$

函数 $f(x)$ 是 $\mathbf{R}$ 上的局部 Lipschitz 函数. 根据广义方向导数和经典方向导数的定义直接计算, 得 $f^\circ(0;\pm1)=1$, 而 $f'(0;\pm1)=0$, 所以 $f^\circ(0;\pm1)\neq f'(0;\pm1)$, 即经典方向导数和广义方向导数在 $x=0$ 处不相等.

定理 4.1.1　设 $f(x)$ 为 $\mathbf{R}^n$ 上的局部 Lipschitz 函数, 在点 $x\in\mathbf{R}^n$ 附近的 Lipschitz 常数为 L, 则有下述结论:

(1) 对固定的 x, 函数 $d\to f^\circ(x;d)$ 是有限次线性的, 且有 $|f^\circ(x;d)|\leqslant L\|d\|$;

(2) $f^\circ(x;d)$ 作为 (x,d) 的函数是上半连续的, 固定 x 作为 d 的函数是局部 Lipschitz 的, 其 Lipschitz 常数为 L;

(3) $f^\circ(x;-d)=(-f)^\circ(x;d),\ \forall d\in\mathbf{R}^n$.

证明　(1) 根据 Lipschitz 条件, 式 (4.1.1) 右端差商当 y 充分接近 x 和 t 充分小时是有界的, 且 $L\|d\|$ 为其上界, 于是 $L\|d\|$ 也是 $f^\circ(x;d)$ 的一个上界, 即 $|f^\circ(x;d)|\leqslant L\|d\|$ 成立.

现在考虑函数 $d\to f^\circ(x;d)$ 的次可加性. 通过计算得

$$\begin{aligned} f^\circ(x;d_1+d_2)&=\limsup_{\substack{y\to x\\ t\to0^+}}\frac{f(y+td_1+td_2)-f(y)}{t}\\ &\leqslant\limsup_{\substack{y\to x\\ t\to0^+}}\frac{f(y+td_1+td_2)-f(y+td_2)}{t}+\limsup_{\substack{y\to x\\ t\to0^+}}\frac{f(y+td_2)-f(y)}{t}.\end{aligned}$$

上式右端第二项为 $f^\circ(x;d_2)$; 对于第一项, 考虑 $y+td_2\to x$, 所以有

$$\begin{aligned}\limsup_{\substack{y\to x\\ t\to0^+}}\frac{f(y+td_1+td_2)-f(y+td_2)}{t}&\leqslant\limsup_{\substack{z\to x\\ t\to0^+}}\frac{f(z+td_1)-f(z)}{t}\\ &=f^\circ(x;d_1),\end{aligned}$$

于是得

$$f^\circ(x;d_1+d_2)\leqslant f^\circ(x;d_1)+f^\circ(x;d_2),$$

故 $d\to f^\circ(x;d)$ 是次可加性的. 根据广义方向导数的定义, $f^\circ(x;\lambda d)=\lambda f^\circ(x;d),\forall\lambda\geqslant 0$ 显然成立, 这说明 $d\to f^\circ(x;d)$ 又是正齐次的, 故 $d\to f^\circ(x;d)$ 是次线性的, 结论 (1) 成立.

(2) 设 $\{x_i\}$ 和 $\{d_i\}$ 分别收敛于 x 和 d 的序列, 对每个 i, 根据上极限定义, 存在 $y_i\in\mathbf{R}^n$ 和 $t_i>0$, 使得

$$\|y_i-x_i\|+t_i<\frac{1}{i},$$

$$f^\circ(x_i;d_i)-\frac{1}{i}\leqslant\frac{f(y_i+t_id_i)-f(y_i)}{t_i}$$
$$=\frac{f(y_i+t_id)-f(y_i)}{t_i}+\frac{f(y_i+t_id_i)-f(y_i+t_id)}{t_i}.$$

根据 $f(x)$ 的 Lipschitz 性质, 上式中第二项不超过 $L\|d_i-d\|$, 于是, 对上式关于 $i\to\infty$ 取上极限得

$$\lim_{i\to\infty}\sup f^\circ(x_i;d_i)\leqslant f^\circ(x;d),$$

故 $f^\circ(x;d)$ 是上半连续的.

根据 $f(x)$ 的 Lipschitz 性质, 当 y 充分接近 x 和 $t>0$ 充分小时, 有

$$f(y+td_1)-f(y)\leqslant f(y+td_2)-f(y)+L\|d_1-d_2\|t.$$

在上式中两边除以 t, 然后关于 $y\to x$ 和 $t\to 0^+$ 取上极限得

$$f^\circ(x;d_1)\leqslant f^\circ(x;d_2)+L\|d_1-d_2\|. \tag{4.1.2}$$

进一步, 在式 (4.1.2) 中 d_1,d_2 互换位置即得函数 $f^\circ(x;d)$ 关于 d 的局部 Lipschitz 性质, 结论 (2) 得证.

(3) 经推导得

$$f^\circ(x;-d)=\lim_{\substack{x'\to x\\ t\to 0^+}}\sup\frac{f(x'-td)-f(x')}{t}$$
$$=\lim_{\substack{y\to x\\ t\to 0^+}}\sup\frac{(-f)(y+td)-(-f)(y)}{t}$$
$$=(-f)^\circ(x;d),\quad y=x'-td$$

结论 (3) 成立. 定理得证.

定理 4.1.1 说明, 局部 Lipschitz 函数的广义方向导数关于方向是次线性函数, 根据 Hahn-Banach 定理, 它可作为一个凸紧集的支撑函数, 这个凸紧集就称为广义梯度.

定义 4.1.2 设 $f(x)$ 为 $\mathbf{R}^n$ 上的局部 Lipschitz 函数, $f(x)$ 的广义梯度, 记为 $\partial f(x)$, 定义如下:

$$\partial f(x)=\left\{\xi\in\mathbf{R}^n|\xi^{\mathrm{T}}d\leqslant f^\circ(x;d),\forall d\in\mathbf{R}^n\right\}. \tag{4.1.3}$$

为区别起见, 广义方向导数和广义梯度有时也分别称为 Clarke 广义方向导数和 Clarke 广义梯度 (或 Clarke 次微分), 广义梯度也可记为 $\partial_{Cl}f(x)$ 或 $\partial_C f(x)$.

定理 4.1.2 设 $f(x)$ 为 $\mathbf{R}^n$ 上的局部 Lipschitz 函数, $f(x)$ 在 x 点附近的 Lipschitz 常数为 L, 则下述结论成立:

(1) $\partial f(x)$ 为 $\mathbf{R}^n$ 上的凸紧集, 且 $\|\xi\| \leqslant L$, $\forall \xi \in \partial f(x)$;

(2) $f^\circ(x;d) = \max\left\{\xi^{\mathrm{T}}d | \xi \in \partial f(x)\right\}, \forall d \in \mathbf{R}^n$; (4.1.4)

(3) $\xi \in \partial f(x)$ 当且仅当 $f^\circ(x;d) \geqslant \xi^{\mathrm{T}}d$, $\forall d \in \mathbf{R}^n$;

(4) 集值映射 $x \to \partial f(x)$ 是上半连续的.

证明 结论 (2) 和结论 (3) 由广义梯度定义即可看出, 我们只证明结论 (1) 和结论 (4).

(1) 根据广义梯度定义和定理 4.1.1, 对任意 $\xi \in \partial f(x)$, 有

$$\xi^{\mathrm{T}}d \leqslant f^\circ(x;d) \leqslant L\|d\|, \quad \forall d \in \mathbf{R}^n. \tag{4.1.5}$$

若存在 $\xi \in \partial f(x)$ 满足 $\|\xi\| > L$, 则选取 $d = \xi$, 此时式 (4.1.5) 不成立, 故 $\|\xi\| \leqslant L$, $\forall \xi \in \partial f(x)$. 另一方面, 由广义梯度定义易见 $\partial f(x)$ 为凸紧集, 结论 (1) 得证.

(4) 我们将证明: 给定 $\varepsilon > 0$, 存在 $\delta > 0$, 使得

$$\partial f(y) \subset \partial f(x) + B(0,\varepsilon). \tag{4.1.6}$$

用反证法. 如果上述事实不成立, 则存在序列 $\{y_i\}$ 满足 $y_i \to x$, $\xi_i \in \partial f(y_i)$, 使得 $\xi_i \notin \partial f(x) + B(0,\varepsilon)$. 根据凸集分离定理, 存在 $d_i \neq 0$, 使得

$$\xi_i^{\mathrm{T}}d_i \geqslant \xi^{\mathrm{T}}d_i, \quad \forall \xi \in \partial f(x) + B(0,\varepsilon),$$

于是有

$$\begin{aligned}\xi_i^{\mathrm{T}}d_i \geqslant &\max\{\xi^{\mathrm{T}}d_i | \xi \in f(x) + B(0,\varepsilon)\}\\ =&f^\circ(x;d_i) + \varepsilon\|d_i\|.\end{aligned}$$

因为 $f^\circ(x;\cdot)$ 是正齐次性的, 我们只需考虑 d_i 满足 $\|d_i\| = 1$. 选取 $\{\xi_i\}$ 和 $\{d_i\}$ 的收敛子列, 不妨设为 $\{\xi_i\}$ 和 $\{d_i\}$ 本身, 令 $\xi_i \to \xi, d_i \to d$, 得

$$\xi^{\mathrm{T}}d \geqslant f^\circ(x;d) + \varepsilon. \tag{4.1.7}$$

另一方面, $f^\circ(y_i;d) \geqslant \xi_i^{\mathrm{T}}d$(因为 $\xi_i \in \partial f(y_i)$), 令 $i \to \infty$, 利用 $f^\circ(x;d)$ 的上半连续性有 $f^\circ(y;d) \geqslant \xi^{\mathrm{T}}d$, 这与式 (4.1.7) 矛盾, 结论 (4) 成立. 定理得证.

定理 4.1.2 的结论 (2) 说明广义方向导数是广义梯度的支撑函数, 因此广义方向导数与广义梯度之间是相互唯一确定的.

定理 4.1.3 设 $f(x), g(x)$ 为 $\mathbf{R}^n$ 上的局部 Lipschitz 函数, λ 为常数, 则有

(1) $\partial(\lambda f(x)) = \lambda\partial f(x)$;

(2) $\partial(f(x) + g(x)) \subset \partial f(x) + \partial g(x)$.

证明 (1) 当 $\lambda \geqslant 0$ 时, 有 $(\lambda f(x))^\circ = \lambda f^\circ(x)$, 根据式 (4.1.1) 给出的广义梯度表达式以及凸紧集与其支撑函数的关系即得结论 (1). 对于 $\lambda < 0$ 情况, 只需考虑

$\lambda=-1$. 因为 $\xi\in\partial(-f(x))$ 当且仅当

$$(-f)^{\circ}(x;d)\geqslant\xi^{\mathrm{T}}d,\quad \forall d\in\mathbf{R}^n.$$

由定理 4.1.1 的结论 (3), $(-f^{\circ})(x;d)=f^{\circ}(x;-d)$, 于是 $\xi\in\partial(-f(x))$ 当且仅当

$$f^{\circ}(x;-d)\geqslant\xi^{\mathrm{T}}d,\quad \forall d\in\mathbf{R}^n,$$

即

$$f^{\circ}(x;-d)\geqslant(-\xi^{\mathrm{T}})(-d),\quad \forall d\in\mathbf{R}^n,$$

这等价于 $-\xi\in\partial f(x)$, 结论 (1) 得证.

(2) 根据广义方向导数定义, 计算得

$$\begin{aligned}(f+g)^{\circ}(x;d)&=\lim_{\substack{y\to x\\t\to0^+}}\sup\frac{f(y+td)+g(y+td)-f(y)-g(y)}{t}\\&\leqslant\lim_{\substack{y\to x\\t\to0^+}}\sup\frac{f(y+td)-f(y)}{t}+\lim_{\substack{y\to x\\t\to0^+}}\sup\frac{g(y+td)-g(y)}{t}\\&=f^{\circ}(x;d)+g^{\circ}(x;d),\quad\forall d\in\mathbf{R}^n,\end{aligned}$$

用广义梯度表示广义方向导数, 得

$$\begin{aligned}\max_{\xi\in\partial(f(x)+g(x))}\xi^{\mathrm{T}}d&\leqslant\max_{\xi\in\partial f(x)}\xi^{\mathrm{T}}d+\max_{\xi\in\partial g(x)}\xi^{\mathrm{T}}d\\&=\max_{\xi\in\partial f(x)+\partial g(x)}\xi^{\mathrm{T}}d,\quad\forall d\in\mathbf{R}^n.\end{aligned}$$

根据凸紧集与其支撑函数的关系, 即得结论 (2). 定理得证.

定理 4.1.3 可以推广到一般形式, 设 $f_i(x),i=1,2,\cdots,m$ 是 $\mathbf{R}^n$ 上的局部 Lipschitz 函数, λ_i, $i=1,\cdots,m$ 为常数, 则有

$$\partial\left(\sum_{i=1}^{m}\lambda_i f_i(x)\right)\subset\sum_{i=1}^{m}\lambda_i\partial f_i(x).\tag{4.1.8}$$

定理 4.1.4 设 $f(x)$ 为 $\mathbf{R}^n$ 上的局部 Lipschitz 函数, 如果 x^* 为 $f(x)$ 的极大(小) 值点, 则 $0\in\partial f(x^*)$.

证明 注意到 $\partial(-f(x))=-\partial f(x)$, 因此只需考虑极小值情况. 由于 x^* 为 $f(x)$ 的极小值点, 则有

$$f^{\circ}(x^*;d)\geqslant0,\quad\forall d\in\mathbf{R}^n$$

(否则存在 $d_1\in\mathbf{R}^n$, 使得 $f^{\circ}(x^*;d_1)<0$, 根据广义方向导数定义易见 x^* 不是 $f(x)$ 的极小值点), 于是根据支撑函数的性质有 $0\in\partial f(x^*)$. 定理得证.

满足 $0\in\partial f(x)$ 的点 x 称为 $f(x)$ 的稳定点 (广义梯度形式下的). 当 $f(x)$ 为光滑函数时, $\partial f(x)=\{\nabla f(x)\}$, 此时 $0\in\partial f(x)$ 等价于 $\nabla f(x)=0$.

4.2 可微性和 Lipschitz 函数的正则性

本节首先介绍几种经典的可微性及有关性质, 然后介绍局部 Lipschitz 函数的正则性, 即广义方向导数与经典方向导数相等.

4.2.1 可微性

如果 $\mathbf{R}^n$ 上函数 $f(x)$ 的方向导数存在, 且存在 $A \in \mathbf{R}^n$, 使得 $f'(x;d) = A^{\mathrm{T}}d$, 则称 $f(x)$ 是 Gâteaux 可微的, A 称为 $f(x)$ 的 Gâteaux 微分, 也称为梯度, 记为 $\nabla f(x) = A$.

$\mathbf{R}^n$ 上函数 $f(x)$ 的 Dini 上、下导数定义如下:

$$f_D^{\uparrow}(x;d) = \limsup_{t\to 0^+} \frac{f(x+td)-f(x)}{t},$$

$$f_D^{\downarrow}(x;d) = \liminf_{t\to 0^+} \frac{f(x+td)-f(x)}{t}.$$

易见, 对于局部 Lipschitz 函数 Dini 上、下导数都存在. 如果 Dini 上、下导数都存在且相等, 即

$$f_D^{\uparrow}(x;d) = f_D^{\downarrow}(x;d),$$

则函数 $f(x)$ 是方向可微的, 且有

$$f'(x;d) = f_D^{\uparrow}(x;d) = f_D^{\downarrow}(x;d).$$

如果 x^* 是 $f(x)$ 的局部极小值点, 则有

$$f_D^{\uparrow}(x^*;d) \geqslant 0, \quad \forall d \in \mathbf{R}^n;$$

如果对任意的 $d \in \mathbf{R}^n$ 都有

$$f_D^{\uparrow}(x^*;d) > 0,$$

则 x^* 是 $f(x)$ 的严格局部极小值点.

$\mathbf{R}^n$ 上函数 $f(x)$ 的 Hadamard 导数 $f'_H(x;d)$ 定义如下:

$$f'_H(x;d) = \lim_{\substack{t\to 0^+\\ d'\to d}} \frac{f(x+td')-f(x)}{t}. \tag{4.2.1}$$

事实上, Hadamard 可微就是一致方向可微, 即下式右端极限

$$f'(x\,;d) = \lim_{t\to 0} \frac{f(x+td)-f(x)}{t}$$

关于 $d \in \mathbf{R}^n$ 一致成立.

在有限维空间中 Hadamard 可微等同与 Fréchet 可微, 在无限维空间中二者则有所区别.

设 $f(x)$ 为 $\mathbf{R}^n$ 上的实函数, 如果存在 $A \in \mathbf{R}^n$, 使得下式成立:

$$\lim_{\substack{y \to x \\ t \to 0^+}} \frac{f(y+td)-f(y)}{t} = A^{\mathrm{T}}d, \quad \forall d \in \mathbf{R}^n,$$

且上式左端极限对于任一个紧集中的 d 是一致成立的, 则称 $f(x)$ 在 x 点是严格可微的, A 称为其微分, 记为 $\nabla f(x) = A$.

命题 4.2.1 设 $f(x)$ 为 $\mathbf{R}^n$ 上的局部 Lipschitz 函数, 且在 x 点是 Gâteaux(Hadamard, 严格, Fréchet) 可微的, 则有 $\nabla f(x) \in \partial f(x)$.

证明 根据可微的定义, 方向导数存在且有 $f'(x;d) = \nabla f(x)^{\mathrm{T}}d$, 于是

$$\nabla f(x)^{\mathrm{T}}d = f'(x;d) \leqslant f^{\circ}(x;d),$$

再根据支撑函数与凸紧集的关系, 有 $\nabla f(x) \in \partial f(x)$. 命题得证.

例 4.2.1 设$f(x) = x^2 \sin\dfrac{1}{x}$, 显然 $f(x)$ 是 $\mathbf{R}$ 上的可微函数, 当然也是局部 Lipschitz 的. 根据广义方向导数定义计算得 $f^{\circ}(0;d) = |d|$, 于是 $\partial f(0) = [-1,1]$, 而 $\nabla f(0) = 0$, 广义梯度不是梯度形成的单点集. 这一现象的原因是 $f(x)$ 在 $x=0$ 点仅仅是可微的, 但在 $x=0$ 点附近不是连续可微的, 是可微但非光滑的.

定理 4.2.1 设 $f(x)$ 为 $\mathbf{R}^n$ 上的局部 Lipschitz 函数, 如果 $f(x)$ 在 x 点是严格可微的, $\nabla f(x)$ 为其梯度, 则有 $\partial f(x) = \{\nabla f(x)\}$; 反过来, 如果 $\partial f(x)$ 是单点集, 记 $\partial f(x) = \{\xi\}$, 则 $f(x)$ 在 x 点是严格可微的, 且有 $\nabla f(x) = \xi$.

证明 假设 $f(x)$ 在 x 点是严格可微的, 根据严格可微及广义方向导数的定义, 易见

$$f^{\circ}(x;d) = \nabla f(x)^{\mathrm{T}}d, \quad \forall d \in \mathbf{R}^n,$$

故

$$\partial f(x) = \{\nabla f(x)\}.$$

假设 $\partial f(x)$ 是单点集, 记 $\partial f(x) = \{\xi\}$, 由广义方向导数与广义梯度的关系 $f^{\circ}(x;d)$ 为集合 $\{\xi\}$ 的支撑函数, 于是

$$f^{\circ}(x;d) = \xi^{\mathrm{T}}d, \quad \forall d \in \mathbf{R}^n.$$

经推导得

$$\begin{aligned}
\liminf_{\substack{y\to x\\ t\to 0^+}} \frac{f(y+td)-f(y)}{t} &= -\limsup_{\substack{y\to x\\ t\to 0^+}} \frac{f(y)-f(y+td)}{t} \\
&= -\limsup_{\substack{y\to x\\ t\to 0^+}} \frac{f(y+td-td)-f(y+td)}{t} \\
&= \limsup_{\substack{y\to x\\ t\to 0^+}} \frac{f(y+td)-f(y)}{t}. \qquad (4.2.2)
\end{aligned}$$

由式 (4.2.2) 知极限

$$\lim_{\substack{y\to x\\ t\to 0^+}} \frac{f(y+td)-f(y)}{t}$$

存在且与 $f^\circ(x;d)$ 相等, 再由 $f^\circ(x;d)=\xi^{\mathrm{T}}d$ 得

$$\lim_{\substack{y\to x\\ t\to 0^+}} \frac{f(y+td)-f(y)}{t} = \xi^{\mathrm{T}}d, \quad \forall d\in\mathbf{R}^n.$$

根据严格可微定义, $f(x)$ 是严格可微的, 且梯度为 $\nabla f(x)=\xi$. 定理得证.

定理 4.2.2 设 $f(x)$ 为 $\mathbf{R}^n$ 上的局部 Lipschitz 函数, $\partial f(x)$ 在点 x 的一个邻域内每一点均为单点集的充要条件是 $f(x)$ 在这个邻域上是连续可微的.

证明 充分性. 假设 $f(x)$ 是连续可微的, 由后面的广义梯度公式 (4.4.1) 易见 $\partial f(x)$ 为单点集.

必要性. 假设 $\partial f(x)$ 均为单点集, 由定理 4.2.1, $f(x)$ 是可微的, 且有 $\partial f(x)=\{\nabla f(x)\}$. 注意到对单点的集值映射上半连续等价于连续, $x\to\partial f(x)$ 的上半连续保性 $\nabla f(x)$ 的连续性. 定理得证.

4.2.2 正则性

定义 4.2.1 设 $f(x)$ 为 $\mathbf{R}^n$ 上的局部 Lipschitz 函数, 如果 $f(x)$ 是方向可微的且广义方向导数与经典方向导数相等, 即 $f'(x;d)=f^\circ(x;d), \forall d\in\mathbf{R}^n$, 则 $f(x)$ 称为是正则的.

定理 4.2.3 设 $f(x)$ 为 $\mathbf{R}^n$ 上的局部 Lipschitz 函数, 则有下述结论:

(1) 如果 $f(x)$ 在 x 处是严格可微的, 则 $f(x)$ 在 x 处是正则的;

(2) 如果 $f(x)$ 是凸函数, 则 $f(x)$ 是正则的;

(3) 如果 $f(x)$ 在 x 点是 Gâteaux 可微且正则的, 则有 $\partial f(x)=\{\nabla f(x)\}$.

证明 (1) $f(x)$ 是严格可微的, 根据定理 4.2.1, $\partial f(x)=\{\nabla f(x)\}$, 于是

$$\begin{aligned}
f^\circ(x;d) &= \max_{\xi\in\{\nabla f(x)\}} \xi^{\mathrm{T}}d = \nabla f(x)^{\mathrm{T}}d \\
&= f'(x;d), \quad \forall d\in\mathbf{R}^n,
\end{aligned}$$

故 $f(x)$ 是正则的.

(2) 由定义不难看出, 广义方向导数可以表述为下述形式

$$f^\circ(x;d) = \lim_{\varepsilon\to 0^+} \sup_{\|y-x\|\leqslant\varepsilon\delta} \sup_{0<t<\varepsilon} \frac{f(y+td)-f(y)}{t},$$

其中 $\delta>0$ 为一固定常数. 根据凸函数性质, 一维函数

$$q(t) = \frac{f(y+td)-f(d)}{t},$$

是非减的, 于是

$$f^\circ(x;d) = \lim_{\varepsilon\to 0} \sup_{\|y-x\|\leqslant\varepsilon\delta} \frac{f(y+\varepsilon d)-f(y)}{\varepsilon}. \tag{4.2.3}$$

根据 $f(x)$ 的 Lipschitz 性质, 对任意 $y\in x+B(0,\varepsilon\delta)$, 有

$$\left|\frac{f(y+\varepsilon d)-f(y)}{\varepsilon} - \frac{f(x+\varepsilon d)-f(x)}{\varepsilon}\right| \leqslant 2\delta L, \tag{4.2.4}$$

其中 L 为 $f(x)$ 的 Lipschitz 常数. 联立式 (4.2.3) 和式 (4.2.4) 得

$$\begin{aligned} f^\circ(x;d) &\leqslant \lim_{\varepsilon\to 0^+} \frac{f(y+\varepsilon d)-f(y)}{\varepsilon} + 2\delta L \\ &= f'(x;d) + 2\delta L. \end{aligned}$$

注意到 δ 的任意性, 由上式得 $f^\circ(x;d)\leqslant f'(x;d)$. 又由于 $f'(x;d)\leqslant f^\circ(x;d)$, 故 $f^\circ(x;d)=f'(x;d)$, $f(x)$ 是正则的.

(3) 由假设条件得

$$f^\circ(x;d) = f'(x;d) = \nabla f(x)^{\mathrm{T}} d, \quad \forall d\in \mathbf{R}^n,$$

再根据支撑函数性质得 $\partial f(x) = \{\nabla f(x)\}$. 定理得证.

推论 4.2.1 设 $f(x)$ 为 $\mathbf{R}^n$ 上的凸函数, 则广义梯度即为凸函数的次微分.

证明 由定理 4.2.3, $f(x)$ 是正则的, 故有

$$\begin{aligned} \max_{\xi\in\partial_{Cl} f(x)} \xi^{\mathrm{T}} d &= f^\circ(x;d) = f'(x;d) \\ &= \max_{\xi\in\partial f(x)} \xi^{\mathrm{T}} d, \quad \forall d\in\mathbf{R}^n, \end{aligned} \tag{4.2.5}$$

上式右端 $\partial f(x)$ 代表凸函数的次微分. 根据凸集与其支撑函数关系, 由式 (4.2.5) 得 $\partial_{Cl} f(x) = \partial f(x)$. 推论得证.

定理 4.2.4 设 $f_1(x),\cdots,f_m(x)$ 为 $\mathbf{R}^n$ 上的局部 Lipschitz 函数, $\lambda_1,\cdots,\lambda_m\geqslant 0$, 如果 $f_i(x)$ 是正则的, 则 $\sum_{i=1}^m \lambda_i f_i(x)_i$ 也是正则的.

证明　利用已知条件及广义方向导数性质, 推导得

$$\begin{aligned}\left(\sum_{i=1}^{m}\lambda_i f_i\right)'(x;d)&=\sum_{i=1}^{m}\lambda_i f_i'(x;d)=\sum_{i=1}^{m}\lambda_i f_i^\circ(x;d)\\&=\sum_{i=1}^{m}(\lambda_i f_i)^\circ(x;d)(\text{因为 }\lambda_i\geqslant 0)\\&\geqslant\left(\sum_{i=1}^{m}\lambda_i f_i\right)^\circ(x;d)(\text{和函数的广义方向导数小于广义方向导数的和})\\&\geqslant\left(\sum_{i=1}^{m}\lambda_i f_i\right)'(x;d),\end{aligned}$$

故

$$\left(\sum_{i=1}^{m}\lambda f_i(x)_i\right)'=\left(\sum_{i=1}^{m}\lambda_i f_i\right)^\circ(x;d),$$

即函数 $\sum\limits_{i=1}^{m}\lambda_i f_i(x)$ 是正则的. 定理得证.

定理 4.2.5　如果 $f_1(x),\cdots,f_m(x)$ 为 $\mathbf{R}^n$ 上的正则局部 Lipschitz 函数, $\lambda_1,\cdots,\lambda_m\geqslant 0$, 则有

$$\partial\left(\sum_{i=1}^{m}\lambda_i f_i(x)\right)=\sum_{i=1}^{m}\lambda_i\partial f_i(x).\tag{4.2.6}$$

证明　根据定理 4.2.4, 函数 $\sum\limits_{i=1}^{m}\lambda_i f_i(x)$ 是正则的, 于是有

$$\begin{aligned}\left(\sum_{i=1}^{m}\lambda_i f_i\right)^\circ(x;d)&=\sum_{i=1}^{m}(\lambda_i f_i)'(x;d)=\sum_{i=1}^{m}\lambda_i f_i'(x;d)\\&=\sum_{i=1}^{m}\lambda_i f_i^\circ(x;d)=\sum_{i=1}^{m}\lambda_i\max_{\xi\in\partial f(x)}\xi^{\mathrm{T}}d=\max_{\xi\in\sum\limits_{i=1}^{m}\lambda_i\partial f(x)}\xi^{\mathrm{T}}d,\end{aligned}$$

由广义梯度定义知式 (4.2.6) 成立. 定理得证.

4.3　中值定理与链锁法则

利用广义梯度也可以得到类似于经典分析中的中值定理, 所不同的是这里的中值定理本质上是一种包含关系而非等式关系.

定理 4.3.1 设 $f(x)$ 为 $\mathbf{R}^n$ 上的局部 Lipschitz 函数, 给定 $x, y \in \mathbf{R}^n$, 则存在 x, y 连线上的点 u 和 $\xi \in \partial f(u)$, 使得

$$f(y) - f(x) = \xi^{\mathrm{T}}(y - x), \tag{4.3.1}$$

结论 (4.3.1) 也可以表述为

$$f(y) - f(x) \in \bigcup_{0<t<1} \{\xi^{\mathrm{T}}(y - x) | \xi \in \partial f(x_t)\}, \tag{4.3.2}$$

其中 $x_t = x + t(y - x)$.

证明 定义一维函数

$$g(t) = f(x + t(y - x)),$$

显然 $g(t)$ 是 $\mathbf{R}$ 上的局部 Lipschitz 函数. 首先证明下式:

$$\partial g(t) \subset \partial f(x + t(y - x))^{\mathrm{T}}(y - x). \tag{4.3.3}$$

式 (4.3.3) 左、右两端均为 $\mathbf{R}$ 上的凸紧集, 我们考虑其对应的支撑函数, 为此只需考虑方向 $d = \pm 1$, 即证明

$$\max\{\xi d | \xi \in \partial g(t)\} \leqslant \max\{\xi d(y - x) | \xi \in \partial f(x + t(y - x))\}, \quad d = \pm 1. \tag{4.3.4}$$

式 (4.3.4) 左端为广义方向导数 $g^\circ(t; d)$, 经推导得

$$\begin{aligned} g^\circ(t; d) &= \lim_{\substack{s \to t \\ \lambda \to 0^+}} \sup \frac{g(s + \lambda d) - g(s)}{\lambda} \\ &= \lim_{\substack{s \to t \\ \lambda \to 0^+}} \sup \frac{f(x + (s + \lambda d)(y - x)) - f(x + s(y - x))}{\lambda} \\ &\leqslant \lim_{\substack{y' \to x + t(y - x) \\ \lambda \to 0^+}} \sup \frac{f(y' + \lambda d(y - x)) - f(y')}{\lambda} \\ &= f^\circ(x + t(y - x); d(y - x)) \\ &= \max\{\xi^{\mathrm{T}} d(y - x) | \xi \in \partial f(x + t(y - x))\}, \quad d = \pm 1, \end{aligned}$$

故式 (4.3.4) 成立.

下面证明式 (4.3.1), 定义 $[0, 1]$ 上的函数 $\theta(t)$ 如下:

$$\theta(t) = f(x + t(y - x)) + t(f(x) - f(y)).$$

注意到, $\theta(0) = \theta(1) = f(x)$, 于是根据 $\theta(t)$ 的连续性知, 存在 $t \in (0, 1)$, 使得 $\theta(t)$ 在 t 处达到极大 (小) 值, 再根据定理 4.1.4 有 $0 \in \partial\theta(t)$. 计算 $\theta(t)$ 的广义梯度得

$$\partial\theta(t) = f(x) - f(y) + \partial f(x + t(y - x))^{\mathrm{T}}(y - x),$$

于是由 $0 \in \partial\theta(t)$ 及上式得式 (4.3.2). 定理得证.

在光滑情形下, 定理 3.2.1 给出的中值定理就是微积分中的 Lagrange 中值定理.

定理 4.3.2 (链锁法则)　设 $g(u)$ 为 $\mathbf{R}^m$ 上的局部 Lipschitz 函数, $F(x)$ 为 $\mathbf{R}^n$ 到 $\mathbf{R}^m$ 上的局部 Lipschitz 函数, 则复合函数 $f(x) = g(F(x))$ 的广义梯度满足

$$\partial f(x) \subset \mathrm{co}\{\partial(\gamma^{\mathrm{T}} F(x)) | \gamma \in \partial g(z)|_{z=F(x)}\}. \tag{4.3.5}$$

证明　为证明式 (4.3.5), 只需证明式 (4.3.5) 右端集合的支撑函数大于 $f(x)$ 的广义方向导数, 为此只需证明对于给定的 $d \in \mathbf{R}^n$, 存在 $\gamma \in \partial g(z)|_{z=F(x)}$(或简记为 $\gamma \in \partial g(F(x))$) 和 $\xi \in \partial(\gamma^{\mathrm{T}} F(x))$, 使得 $f^\circ(x;d) \leqslant \xi^{\mathrm{T}} d$. 根据广义方向导数定义, 存在 $y_k \to x, t_k \to 0^+$, 使得

$$\lim_{k \to \infty} \frac{f(y_k + t_k d) - f(y_k)}{t_k} = f^\circ(x, d),$$

再根据定理 4.3.1 的广义梯度形式中值定理, 对每个 k, 存在 $F(y_k)$ 和 $F(y_k + t_k d)$ 连线上的点 z_k 和 $\gamma_k \in \partial g(z_k)$, 使得

$$\begin{aligned}\frac{f(y_k + t_k d) - f(y_k)}{t_k} &= \frac{g(F(y_k + t_k d)) - g(F(y_k))}{t_k} \\ &= \gamma_k^{\mathrm{T}} \frac{F(y_k + t_k d) - F(y_k)}{t_k}.\end{aligned}$$

注意到有 $z_k \to F(x)$, 则存在 $\{\gamma_k\}$ 的子序列, 不妨设为 $\{\gamma_k\}$ 本身, 使得 $\gamma_k \to \gamma \in \partial g(F(x))$, 此 γ 正为我们所需.

下面证明 ξ 的存在性, 根据广义梯度形式中值定理, 存在 y_k 与 $y_k + t_k d$ 连线上的 w_k, $\xi_k \in \partial(\gamma^{\mathrm{T}} F(x))|_{x=w_k}$, 使得

$$\gamma_k^{\mathrm{T}} \frac{F(y_k + t_k d) - F(y_k)}{t_k} = \xi_k^{\mathrm{T}} d.$$

由于 $w_k \to x$, 而序列 $\{\xi_k\}$ 是有界的, 故序列 $\{\xi_k^{\mathrm{T}} d\}$ 也是有界的. 我们可选取 $\{\xi_k^{\mathrm{T}} d\}$ 中的一个收敛子列, 不妨记为 $\{\xi_k^{\mathrm{T}} d\}$ 本身, 这样 $\xi_k^{\mathrm{T}} d \to \xi^{\mathrm{T}} d$, 根据广义梯度的上半连续性, 有 $\xi \in \partial(\gamma^{\mathrm{T}} F(x))$.

综合以上得

$$\begin{aligned}\frac{f(y_k + t_k d) - f(y_k)}{t_k} &= ((\gamma_k - \gamma)\gamma)^{\mathrm{T}} \frac{F(y_k + t_k d) - F(y_k)}{t_k} \\ &= (\gamma_k - \gamma)^{\mathrm{T}} \frac{F(y_k + t_k d) - F(y_k)}{t_k} + \xi_k^{\mathrm{T}} d.\end{aligned}$$

由于 $F(x)$ 是 Lipschitz 连续的, 所以

$$\frac{F(y_k + t_k d) - F(y_k)}{t_k}$$

是有界的, 另由于 $\gamma_k \to \gamma$, 于是, 取极限得

$$f^\circ(x; d) = \lim_{k\to\infty} \frac{f(y_k + t_k d) - f(y_k)}{t_k} = \xi^{\mathrm{T}} d.$$

上式的 ξ 正为我们所求. 定理得证.

注 4.3.1 在定理 4.3.2 中记 $F(x) = (f_1(x), \cdots, f_m(x))^{\mathrm{T}}$, 则由式 (4.3.5) 可得

$$\partial f(x) \subset \mathrm{co}\left\{ \sum_{i=1}^{n} r_i \xi_i | \xi_i \in \partial h_i(x), (r_i, \cdots, r_m) \in \partial g(F(x)) \right\}.$$

利用定理 4.3.2, 我们可以得到两个局部 Lipschitz 函数乘积的广义梯度性质, 设 $f(x)$ 和 $g(x)$ 为 $\mathbf{R}^n$ 上的局部 Lipschitz 函数, 则有

$$\partial(f(x)g(x)) \subset f(x)\partial g(x) + g(x)\partial f(x).$$

例 4.3.1 考虑极大值函数

$$f(x) = \max_{1\leqslant i\leqslant m} f_i(x), \tag{4.3.6}$$

其中每个 $f_i(x)$ 为 $\mathbf{R}^n$ 上的局部 Lipschitz 函数. 根据命题 4.1.1, $f(x)$ 也是局部 Lipschitz 函数, 以下讨论 $f(x)$ 的广义梯度. 定义 $\mathbf{R}^m$ 上的函数 $g(z)$ 如下:

$$g(z) = g(z_1, \cdots, z_m) = \max\{z_i | i = 1, \cdots, m\},$$

记

$$F(x) = (f_1(x), \cdots, f_m(x)),$$

这样, 式 (4.3.6) 所给出的函数 $f(x)$ 可表示为 $f(x) = g(F(x))$. 注意到, $g(z)$ 为线性函数的极大值函数, 因此是凸函数, 根据推论 3.3.2 得

$$\partial g(z) = \left\{ \xi_1, \cdots, \xi_m | \xi_i \geqslant 0, \sum_{i=1}^{m} \xi_i = 1,\ \text{如果 } i \notin I(z),\ \text{则 } \xi_i = 0 \right\},$$

其中 $I(z) = \{i | z_i = \max_{1\leqslant k\leqslant m} z_k\}$. 由于凸函数广义梯度即为凸函数的次微分, 利用链锁法则得

$$\partial f(x) \subset \mathrm{co}\{\partial f_i(x) | i \in I(x)\},$$

其中 $I(x) = \{i | f_i(x) = f(x)\}$.

4.4　广义梯度公式及广义 Jacobi

本节讨论向量值局部 Lipschitz 函数的广义 Jacobi. 在介绍广义 Jacobi 之前我们先讨论广义梯度的另外一种表示. 我们知道局部 Lipschitz 函数是几乎处处可微的 (Rademacher 定理), 即去掉一个测度为零的集合以外全是可微的.

4.4.1　广义梯度公式

定理 4.4.1 (广义梯度公式)　设 $f(x)$ 为 $\mathbf{R}^n$ 上的局部 Lipschitz 函数, 集合 $\Omega \subset \mathbf{R}^n$ 测度为零, 记 Ω_f 为 $\mathbf{R}^n$ 中所有 $f(x)$ 不可微点组成的集合, 则广义梯度有下述表达式:

$$\partial f(x) = \mathrm{co}\{\lim_{i\to\infty} \nabla f(x_i) | x_i \to x, x_i \notin \Omega \cup \Omega_f\}. \tag{4.4.1}$$

证明　式 (4.4.1) 右端的含意为: 对于序列 $\{x_i\}$, 其中每个 x_i 不属于 $\Omega \cup \Omega_f$, 而 x_i 收敛于 x, $\nabla f(x_i)$ 也有极限. 因为 $f(x)$ 几乎处处可微, 所以 $\Omega \cup \Omega_f$ 是测度为零的集合, 另一方面, $f(x)$ 的 Lipschitz 性质可保证在 x 点附近 $\nabla f(x_i)$ 有界, 因此, 存在使 $\{\nabla f(x_i)\}$ 存在且收敛的 x_i, 故式 (4.4.1) 右端是有意义的, 并且是非空的. 根据广义梯度的上半连续性及 $\nabla f_i(x_i) \in \partial f(x_i)$, 可知 $\lim\limits_{x_i\to x} \nabla f_i(x_i) \in \partial f(x_i)$, 于是

$$\left\{\lim \nabla f_i(x_i) | x_i \to x, x_i \notin \Omega \bigcup \Omega_f\right\} \subset \partial f(x).$$

再根据 $\partial f(x)$ 的凸性, 得

$$\mathrm{co}\left\{\lim \nabla f(x_i) | x_i \to x, x_i \notin \Omega \bigcup \Omega_f\right\} \subset \partial f(x).$$

因此, 为证明式 (4.4.1), 只需证明

$$\partial f(x) \subset \mathrm{co}\left\{\lim \nabla f_i(x_i) | x_i \to x, x_i \notin \Omega \bigcup \Omega_f\right\},$$

或等价地, 证明 $f^\circ(x;d)$ 小于式 (4.4.1) 右端集合的支持函数. 因此, 我们只要证明

$$f^\circ(x;d) \leqslant \limsup\left\{\nabla f(y)^{\mathrm{T}} d | \, y \to x, y \notin \Omega \bigcup \Omega_f\right\}. \tag{4.4.2}$$

以下证明, 对任意 $d \in \mathbf{R}^n$, 且 $d \neq 0$ 和任意的 $\varepsilon > 0$, 有

$$f^\circ(x;d) - \varepsilon \leqslant \limsup\left\{\nabla f(y)^{\mathrm{T}} d | \, y \to x, y \notin \Omega \bigcup \Omega_f\right\}. \tag{4.4.3}$$

记式 (4.4.3) 左端项为 α, 由广义梯度的上半连续性及 $\nabla f(y) \in \partial f(y)$, 可选取 $\delta > 0$, 使得当

$$y \in x + B(0,\delta),\ y \notin \Omega \bigcup \Omega_f,$$

有

$$\nabla f(y)^{\mathrm{T}} d \leqslant \alpha+\varepsilon.$$

考虑线段

$$L_y=\left\{y+td \mid 0<t<\frac{\delta}{2\|d\|}\right\},$$

因为 $\Omega \bigcup \Omega_f$ 测度为零, 根据 Fubini 定理, 几乎对所有的 $y \in x+B\left(0, \frac{\delta}{2}\right)$, L_y 与 $\Omega \bigcup \Omega_f$ 相交于一个一维零测度集. 设 y 是具有此特征的且在 $x+B\left(0, \frac{\delta}{2}\right)$ 内的点, 令 $t \in\left(0, \frac{\delta}{2\|d\|}\right)$, 由于 $\nabla f(y)$ 在 L_y 上几乎处处可微, 则有

$$f(y+td)-f(y)=\int_0^t \nabla f(y+sd)^{\mathrm{T}} d \mathrm{d} s.$$

又由于

$$\|y+sd-x\|<\delta, \quad \forall 0<s<t,$$

于是

$$\nabla f(y+sd)^{\mathrm{T}} d \leqslant \alpha+\varepsilon,$$

进而有

$$f(y+td)-f(y) \leqslant t(\alpha+\varepsilon). \tag{4.4.4}$$

注意到式 (4.4.4)对几乎所有的 $y \in x+B\left(0, \frac{\delta}{2}\right)$ 和所有的 $t \in\left(0, \frac{\delta}{2\|d\|}\right)$ 都成立, 又 $f(x)$ 是连续的, 故式 (4.4.4) 对所有的 $y \in x+B\left(0, \frac{\delta}{2}\right)$ 和所有的 $t \in\left(0, \frac{\delta}{2\|d\|}\right)$ 都成立. 根据广义方向导数定义, 由式 (4.4.4) 得 $f^{\circ}(x ; d) \leqslant \delta+\varepsilon$, 即式 (4.4.3) 成立. 定理得证.

公式 (4.4.1) 也可作为广义梯度的等价定义, 该定义较定义 4.1.2 更为直观, 并且可应用于广义梯度的具体计算. 在无限维空间中则有所不同, 此时广义梯度没有式 (4.4.1) 形式的表达式.

推论 4.4.1 设 $f(x)$ 为 $\mathbf{R}^n$ 上局部 Lipschitz 函数, 则有

$$f^{\circ}(x ; d)=\limsup_{y \rightarrow x}\left\{\nabla f(y)^{\mathrm{T}} d \mid y \notin \Omega \bigcup \Omega_f\right\},$$

其中, Ω, Ω_f 定义同定理 4.4.1.

例 4.4.1 设

$$f(x, y)=\max \{\min \{x,-y\}, y-x\}.$$

记

$$S_1 = \{(x,y)|y \leqslant 2x, y \leqslant -x\},$$

$$S_2 = \left\{(x,y)|y \leqslant \frac{1}{2}x, y \geqslant -x\right\},$$

$$S_3 = \left\{(x,y)|y \leqslant 2x \text{ 或 } y \geqslant \frac{1}{2}x\right\},$$

易见 $S_1 \bigcup S_2 \bigcup S_3 = \mathbf{R}^2$, 且 $f(x,y)$ 可表示为

$$f(x,y) = \begin{cases} x, & (x,y) \in S_1, \\ -y, & (x,y) \in S_2, \\ y-x, & (x,y) \in S_3. \end{cases}$$

注意到, 集合 S_1, S_2, S_3 的边界组成的集合 Ω 测度为零, 而当 $(x,y) \notin \Omega$ 时 $f(x,y)$ 是可微的, 其梯度分别是 $(1,0)$, $(0,-1)$, $(-1,1)$, 于是, 根据定理 4.4.1 得

$$\partial f(0,0) = \mathrm{co}\{(1,0),(0,-1),(-1,1)\}.$$

4.4.2 广义 Jacobi

设 $F(x)$ 为 $\mathbf{R}^n$ 到 $\mathbf{R}^m$ 上的 D_1 局部 Lipschitz 函数, 记 $F(x) = (f_1(x),\cdots, f_m(x))^{\mathrm{T}}$, 其中每个 $f_i(x)$ 为 $\mathbf{R}^n$ 上的局部 Lipschitz 函数. 根据 Rademacher 定理, 每个 $f_i(x)$ 是几乎处处可微的, 因此, $F(x)$ 也是几乎处处可微的, 即 $JF(x)$ 几乎处处存在. 借用这一事实, 我们可将广义梯度概念推广到向量值函数, 采用的是定理 4.4.1 给出的广义梯度公式, 而不是利用广义方向导数.

定义 4.4.1 设 $F(x) = (f_1(x),\cdots,f_m(x))$, 其中 $f_i(x)$, $i=1,\cdots,m$ 为 $\mathbf{R}^n$ 上的局部 Lipschitz 函数, 记 Ω_F 为 $F(x)$ 不可微点形成的集合, 广义 Jacobi $\partial F(x)$ 定义如下:

$$\partial F(x) = \mathrm{co}\{\lim JF(x_i)|x_i \to x, x_i \notin \Omega_F\}. \tag{4.5.5}$$

广义 Jacobi 通常也称为 Clarke 广义 Jacobi, 从定义可以看出, 广义 Jacobi 为 $m \times n$ 阶矩阵空间中的一个凸集.

下述定理给出了广义 Jacobi 的一些性质.

定理 4.4.2 设 $F(x)$ 为 $\mathbf{R}^n$ 到 $\mathbf{R}^m$ 上的局部 Lipschitz 函数, 记 $F(x) = (f_1(x),\cdots,f_m(x))$, 则有

(1) $\partial F(x)$ 是 $\mathbf{R}^{m\times n}$ 空间中的凸紧集;

(2) 集值映射 $x \to \partial F(x)$ 是上半连续的, 即对给定 $x \in \mathbf{R}^n$, 对任意 $\varepsilon > 0$, 存在 $\delta > 0$, 使得下式成立:

$$\partial F(y) \subset \partial F(x) + \varepsilon B_{m\times n}(0,1), \quad \forall y \in x + \varepsilon B_{m\times n}(0,1),$$

其中 $B_{m\times n}(0,1)$ 为 $\mathbf{R}^{m\times n}$ 空间中的单位球;

(3) 记 L_i 为 $f_i(x)$ 在 x 点附近的 Lipschitz 常数, 则 $L=\|(L_1,\cdots,L_m)\|$ 为 $F(x)$ 在 x 点附近的 Lipschitz 常数, 且有 $\partial F(x)\subset LB_{m\times n}(0,1)$;

(4) $\partial F(x)\subset \partial f_1(x)\times\cdots\times\partial f_m(x)$.

证明类似于广义梯度相应的结论, 故从略.

式 (4.4.5) 中的广义 Jacobi 中去掉凸包, 则相应的微分称为 B 微分, 记为 $\partial_B F(x)$, 即

$$\partial_B F(x)=\{\lim JF(x_i)|x_i\to x, x_i\in\Omega_F\}.$$

显然, B 微分不再是凸集, 但它还满足上半连续等性质.

4.5 极大值函数广义 Jacobi 的计算

本节给出计算有限个光滑函数极大值的广义 Jacobi 的方法.

4.5.1 一般形式极大值函数

考虑向量形式的极大值函数:

$$F(x)=(\max_{j\in J_1} f_{1j}(x),\cdots,\max_{j\in J_m} f_{mj}(x))^{\mathrm{T}}, \tag{4.5.1}$$

其中 $J_i, i=1,\cdots,m$ 是有限的指标集, $f_{ij}(x)$ $j\in J_i$, $i=1,\cdots,m$ 是 $\mathbf{R}^n$ 上连续可微函数. 我们将讨论计算函数 $F(x)$ 在一点 x 处的 B 微分和 Clarke 广义 Jacobi 中的一个元素. 为此, 假设 $f_{ij}(x)$ 及 $\nabla f_{ij}(x)$, $j\in J_i$, $i=1,\cdots,m$ 的值是可计算的.

引理 4.5.1 令 $h(x)=\max_{i\in I} h_i(x)$, 其中 $h_i(x)$ 为 $\mathbf{R}^n$ 上的连续可微函数, I 是个有限的指标集. 记

$$I(x)=\{i\in I|h_i(x)=h(x)\},$$

那么 $h(x)$ 在 x 点可微的充要条件是

$$\nabla h_i(x)=\nabla h_j(x),\quad \forall i,j\in I(x).$$

特别地, 如果 $I(x)$ 只包含一个元素, 记 $I(x)=\{i_0\}$, 那么 $h(x)$ 在 x 点可微且有 $\nabla h(x)=\nabla h_{i_0}(x)$.

引理 4.5.2 设 A 是 $m\times n$ 阶矩阵, 那么线性不等式系统 $Ay<0, y\in\mathbf{R}^n$ 的解集是 $\mathbf{R}^n$ 中的一个开凸锥.

引理 4.5.3 假设 $a_i\in\mathbf{R}^n$, $i\in I$, I 是有限指标集且满足 $a_j\neq a_k, \forall j,k\in I, j\neq k$, 如果令 $i_0\in I$ 满足 $\|a_{i_0}\|=\max\limits_{i\in I}\|a_i\|$, 则下式成立:

$$a_i^{\mathrm{T}}a_{i_0}<a_{i_0}^{\mathrm{T}}a_{i_0},\quad \forall i\in I\backslash\{i_0\}.$$

记

$$f_i(x) = \max_{j \in J_i} f_{ij}(x), \quad i = 1, \cdots, m, \tag{4.5.2a}$$

$$J_i(x) = \{j \in J_i | f_{ij}(x) = f_i(x)\}, \quad i = 1, \cdots, m, \tag{4.5.2b}$$

显然 $f_i(x)$ 是方向可微的, 其方向导数为

$$f_i'(x; y) = \max_{j \in J_i(x)} f_{ij}'(x; y), \quad \forall y \in \mathbf{R}^n, \quad i = 1, \cdots, m. \tag{4.5.3}$$

基于指标集 $J_i(x)$, 引入指标集 $\bar{J}_i(x)$, $i = 1, \cdots, m$, 满足下述条件:

$$\bar{J}_i(x) \subset J_i(x), \quad i = 1, \cdots, m, \tag{4.5.4a}$$

$$\nabla f_{ij}(x) \neq \nabla f_{ik}(x), \quad \forall j, k \in \bar{J}_i(x), \quad j \neq k, \quad i = 1, \cdots, m. \tag{4.5.4b}$$

对任意 $t_i \in J_i(x)$, 存在 $k_i \in \bar{J}_i(x)$, 使得

$$\nabla f_{it_i}(x) = \nabla f_{ik_i}(x), \quad i = 1, \cdots, m. \tag{4.5.4c}$$

显然, 指标集 $\bar{J}_i(x)$ 是 $J_i(x)$ 的一个子集, 事实上, 可以通过删除指标集 $J_i(x)$ 中对应满足 $\nabla f_{ij}(x) = \nabla f_{ik}(x)$ 的多余指标而得到. 易见, $f_i'(x; y)$ 可以表示为

$$f_i'(x; y) = \max_{j \in \bar{J}_i(x)} \nabla f_{ij}(x)^{\mathrm{T}} y, \quad \forall y \in \mathbf{R}^n, \quad i = 1, \cdots, m. \tag{4.5.5}$$

显然, 对于固定点 $x \in \mathbf{R}^n$, $f_i'(x; \cdot)$ 是分片光滑函数, 特别是分片仿射函数, 因此也是局部 Lipschitz 的. 根据分片光滑函数性质, 下述关系式成立:

$$\partial_B F'(x; y)|_{y=0} \subset \partial_B F(x), \tag{4.5.6}$$

其中

$$F'(x; y) = (f_1'(x; y), \cdots, f_m'(x; y))^{\mathrm{T}}.$$

给出指标集 $j_i \in \bar{J}_i(x), i = 1, \cdots, m$, 构造下列线性不等式系统:

$$L_{j_1, \cdots, j_m}(x) \ (\nabla f_{ik_i}(x) - \nabla f_{ij_i}(x))^{\mathrm{T}} y < 0, \quad y \in \mathbf{R}^n,$$
$$\forall k_i \in \bar{J}_i(x) \backslash \{j_i\}, \quad i = 1, \cdots, m,$$

或记为

$$\begin{aligned} L_{j_1, \cdots, j_m}(x) \quad & (\nabla f_{1k_1}(x) - \nabla f_{1j_1}(x))^{\mathrm{T}} y < 0, \quad y \in \mathbf{R}^n, \quad \forall k_1 \in \bar{J}_1(x) \backslash \{j_1\}, \\ & (\nabla f_{2k_2}(x) - \nabla f_{2j_2}(x))^{\mathrm{T}} y < 0, \quad y \in \mathbf{R}^n, \quad \forall k_2 \in \bar{J}_2(x) \backslash \{j_2\}, \\ & \cdots\cdots\cdots\cdots \\ & (\nabla f_{mk_m}(x) - \nabla f_{mj_m}(x))^{\mathrm{T}} y < 0, \quad y \in \mathbf{R}^n, \quad \forall k_m \in \bar{J}_m(x) \backslash \{j_m\}. \end{aligned}$$

容易看出, 系统 $L_{j_1,\cdots,j_m}(x)$ 含有 n 个变量和 m 个严格线性不等式组.

定理 4.5.1 假设存在指标集 $j_i \in \bar{J}_1(x), \cdots, j_m \in \bar{J}_m(x)$, 使得它们对应的线性不等式系统 $L_{j_1,\cdots,j_m}(x)$ 是相容 (有解) 的, 那么

$$J(f_{1j_1}(x), \cdots, f_{mj_m}(x))^{\mathrm{T}} \in \partial_B F(x).$$

证明 令 $\bar{y} \in \mathbf{R}^n$ 为系统 $L_{j_1,\cdots,j_m}(x)$ 的解, 那么有

$$\begin{aligned}\nabla f'_{ik_i}(x;\bar{y}) &= \nabla f_{ik_i}(x)^{\mathrm{T}}\bar{y} < \nabla f_{ij_i}(x)^{\mathrm{T}}\bar{y} \\ &< f'_{ij_i}(x;\bar{y}), \forall k_i \in \bar{J}_i(x)\backslash\{j_i\}, \quad i=1,\cdots,m.\end{aligned} \tag{4.5.7}$$

于是, 当 $\lambda > 0$ 时, 有

$$f'_{ik_i}(x;\lambda\bar{y}) < f'_{ij_i}(x;\lambda\bar{y}), \quad \forall k_i \in \bar{J}_i(x)\backslash\{j_i\}, \quad i=1,\cdots,m. \tag{4.5.8}$$

由式 (4.5.5) 和式 (4.5.8) 及引理 4.5.1, 在射线 $y=\lambda\bar{y}, \lambda>0$ 上有

$$\nabla_y f'_i(x;y)|_{y=\lambda\bar{y}} = \nabla_y(f_{ij_i}(x)^{\mathrm{T}}y)|_{y=\lambda\bar{y}} = \nabla f_{ij_i}(x).$$

而每一个 $f'_i(x;\cdot)$ 都是可微的, 由前面所述及 B 微分的定义, 直接推导得

$$\begin{aligned}\xi &= \lim_{\lambda\to 0^+} J_y F'(x;y)|_{y=\lambda\bar{y}} \\ &= \lim_{\lambda\to 0^+} \left(\nabla_y f'_1(x;y)|_{y=\lambda\bar{y}}, \cdots, \nabla_y f'_m(x;y)|_{y=\lambda\bar{y}}\right)^{\mathrm{T}} \\ &= \lim_{\lambda\to 0^+} (\nabla f_{1j_1}(x), \cdots, \nabla f_{mj_m}(x))^{\mathrm{T}} \\ &= J(f_{1j_1}(x), \cdots, f_{mj_m}(x))^{\mathrm{T}} \\ &\in \partial_B F'(x;y)|_{y=0}.\end{aligned} \tag{4.5.9}$$

由式 (4.5.6) 和式 (4.5.9) 得

$$\xi = J(f_{1j_1}(x), \cdots, f_{mj_m}(x))^{\mathrm{T}} \in \partial_B F(x).$$

定理得证.

由定理 4.5.1 知, 计算 $\partial_B F(x)$ 的元素可以转化为寻找指标集

$$j_1 \in \bar{J}_1(x), \cdots, j_m \in \bar{J}_m(x),$$

使得它所对应的系统 $L_{j_1,\cdots,j_m}(x)$ 是相容的. 事实上, 如果对每一组指标

$$j_1 \in \bar{J}_1(x), \cdots, j_m \in \bar{J}_m(x)$$

确定系统 $L_{j_1,\cdots,j_m}(x)$ 的相容性, 则至少可以找到一个相容的系统.

4.5.2　线性函数的极大值

对于一般形式极大值函数, 我们只能计算到 B 微分中的一个元素, 以下考虑特殊情况, 即 $f_{ij}(x)$ 为线性函数, 我们将给出其 B 微分集合的表达式. 在式 (4.5.1) 中假设

$$f_{ij}(x) = a_{ij}^{\mathrm{T}} x + b_{ij}, \quad j \in J_i, \quad i = 1, \cdots, m,$$

其中 $a_{ij} \in \mathbf{R}^n$, $b_{ij} \in \mathbf{R}^1$. 首先给出集合 $\partial_B F(x)$ 的表达式, 见下面的式 (4.5.12), 这样, 可以通过确定几个线性不等式系统的相容性来计算集合 $\partial_B F(x)$. 给定一个点 $x \in \mathbf{R}^n$, 存在 $\delta > 0$, 使得

$$J_i(x') \subset J_i(x), \quad \forall x' \in B(x, \delta),$$

那么 $B(x, \delta)$ 中的 $f_i(x)$, $i = 1, \cdots, m$ 可以写成

$$f_i(x') = \max_{j \in J_i(x)} f_{ij}(x'), \quad \forall x' \in B(x, \delta), \quad i = 1, \cdots, m. \tag{4.5.10}$$

既然 $f_{ij}(x)$, $j \in J_i$, $i = 1, \cdots, m$ 本身是线性的, 不难看出, 在 $B(x, \delta)$ 中, $f_i(x)$, $i = 1, \cdots, m$ 可以表示为

$$f_i(x') = \max_{j \in \bar{J}_i(x)} f_{ij}(x'), \quad \forall x' \in B(x, \delta), \quad i = 1, \cdots, m. \tag{4.5.11}$$

定理 4.5.2　设 $f_{ij}(x)$, $j \in J_i$, $i = 1, \cdots, m$ 为线性函数, 那么

$$\begin{aligned}\partial_B F(x) = \{ & J(f_{1j_1}(x), \cdots, f_{mj_m}(x))^{\mathrm{T}} | L_{j_1, \cdots, j_m}(x) \text{ 相容}, \\ & j_1 \in \bar{J}_1(x), \cdots, j_m \in \bar{J}_m(x)\}.\end{aligned} \tag{4.5.12}$$

证明　利用定理 4.5.1 证明式 (4.5.12). 给定 $\xi \in \partial_B F(x)$, 由 B 微分的定义, 存在一个序列 $\{x_n\}_1^\infty$, 使得 $x_n \to x$, $x_n \in B(x, \delta)$, $F(x)$ 在点 x_n 可微并且 $JF(x_n) \to \xi$. 注意到 $f_i(x)$ 的表达式 (4.5.11) 及

$$\nabla f_{ij}(x) \neq \nabla f_{ik}(x), \quad \forall j, k \in \bar{J}_i(x), \quad j \neq k, \quad i = 1, \cdots, m,$$

由引理 4.5.1, 存在指标 $j_1(x_n) \in \bar{J}_1(x_n), \cdots, j_m(x_n) \in \bar{J}_m(x_n)$, 使得

$$JF(x_n) = J(f_{1j_1(x_n)}(x_n), \cdots, f_{mj_m(x_n)}(x_n))^{\mathrm{T}} \to \xi, \quad x_n \to x. \tag{4.5.13}$$

既然 $\nabla f_{ij}(x)$ 是常数且

$$\nabla f_{ij}(x) \neq \nabla f_{ik}(x), \quad \forall j, k \in \bar{J}_i(x), \quad j \neq k, \quad i = 1, \cdots, m,$$

则存在一组指标

$$j_1(x) \in \bar{J}_1(x), \cdots, j_m(x) \in \bar{J}_m(x)$$

及整数 N, 使得

$$j_1 = j_1(x_n), \cdots, j_m = j_m(x_n), \quad \forall n \geqslant N,$$

也就是说, 存在一个序列 $\{x_n\}_1^\infty$, 使得 $F(x)$ 在点 x_n 是可微的, 且有

$$JF(x_n) = J(f_{1j_1(x_n)}(x_n), \cdots, f_{mj_m(x_n)}(x_n))^{\mathrm{T}}, \quad \forall n \geqslant N.$$

因此, 结合引理 4.5.1 有

$$f_{ik_i}(x_n) < f_{ij_i}(x_n), \quad \forall k_i \in \bar{J}_i(x) \backslash \{j_i\}, \quad i = 1, \cdots, m, \quad n \geqslant N, \tag{4.5.14}$$

即

$$a_{ik_i}^{\mathrm{T}} x_n + b_{ik_i} < a_{ij_i}^{\mathrm{T}} x_n + b_{ij_i}, \quad \forall k_i \in \bar{J}_i(x) \backslash \{j_i\}, \quad i = 1, \cdots, m, \quad n \geqslant N. \tag{4.5.15}$$

注意到

$$a_{ik_i}^{\mathrm{T}} x_n + b_{ik_i} = a_{ij_i}^{\mathrm{T}} x_n + b_{ij_i}, \quad \forall k_i \in \bar{J}_i(x) \backslash \{j_i\}, \quad i = 1, \cdots, m. \tag{4.5.16}$$

由式 (4.5.15) 和式 (4.5.16) 得

$$a_{ik_i}^{\mathrm{T}} (x_n - x) < a_{ij_i}^{\mathrm{T}} (x_n - x), \quad \forall k_i \in \bar{J}_i(x) \backslash \{j_i\}, \quad i = 1, \cdots, m, \tag{4.5.17}$$

即

$$\nabla f_{ik_i}(x)^{\mathrm{T}} (x_n - x) < \nabla f_{ij_i}(x)^{\mathrm{T}} (x_n - x), \quad \forall k_i \in \bar{J}_i(x) \backslash \{j_i\}, \quad i = 1, \cdots, m. \tag{4.5.18}$$

这意味着 $\lim\limits_{n\to\infty} x_n = x$ 是 $L_{j_1,\cdots,j_m}(x)$ 的解, 因此式 (4.5.12) 成立. 定理得证.

基于定理 4.5.2, 当 $f_{ij}(x)$ 是线性函数时, 我们通过确定系统 $L_{j_1,\cdots,j_m}(x)$ 的相容性来计算其 B 微分集合 $\partial_B F(x)$. 由下面的命题我们知道, 确定一个线性不等式系统的相容性可以转为求一个互补线性规划, 因此, 计算集合 $\partial_B F(x)$ 是可实现的.

命题 4.5.1 令 A 为一个 $m \times n$ 矩阵. 线性系统 $Ay < 0, y \in \mathbf{R}^n$ 是相容的当且仅当线性规划问题

$$\begin{aligned}
\min \quad & \sum_{j=1}^{n} z_j, \\
\text{s.t.} \quad & A^{\mathrm{T}} p + (z_1, \cdots, z_n)^{\mathrm{T}} \geqslant 0, \\
& \sum_{i=1}^{m} p_i = 1, \quad p_i \geqslant 0, \quad i = 1, \cdots, m, \quad z_j \geqslant 0, \quad j = 1, \cdots, n
\end{aligned}$$

的目标函数最小值不为零, 其中 p_i 为 p 的第 i 个元素.

4.5.3 复合极大值函数

考虑下述形式的极大值复合函数

$$G(x) = g(\max_{j\in J_1} f_{1j}(x), \cdots, \max_{j\in J_m} f_{mj}(x)), \tag{4.5.19}$$

其中 $f_{ij}(x), j \in J_i, i = 1, \cdots, m$ 和 $g(x)$ 分别是 $\mathbf{R}^n$ 和 $\mathbf{R}^m$ 上的连续可微函数, $J_i, i = 1, \cdots, m$ 为有限指标集. 显然, $G(x)$ 可以写成 $G(x) = G(F(x))$, $F(x)$ 的表达式见 (4.5.1). 利用前面的方法, 基于 B 微分的链锁法则, 我们可以通过计算 $\partial_B F(x)$ 的元素来计算 $G(x)$ 的 B 微分元素. 事实上, 如果 $J(f_{1j_1}(x), \cdots, f_{mj_m}(x))^{\mathrm{T}}$ 是 $\partial_B F(x)$ 的一个元素, 那么 $\nabla g(f_{1j_1}(x), \cdots, f_{mj_m}(x))$ 是 $\partial_B G(x)$ 的一个元素, 也是广义梯度 $\partial G(x)$ 的一个元素.

我们也可以计算由极大值复合函数组成的向量函数的 B 微分元素. 考虑函数

$$\bar{G}(x) = (G_1(x), \cdots, G_p(x))^{\mathrm{T}},$$

其中

$$G_k(x) = g_k(\max_{j\in J_{1k}} f_{1jk}(x), \cdots, \max_{j\in J_{mk}} f_{mjk}(x)), \quad k = 1, \cdots, p,$$

$g_k(x)$ 和 $f_{ijk}(x)$ 是连续可微函数, 每个 J_{ik} 是一个有限指标集. 通过使用链锁法则和计算下列向量值函数的 B 微分元素来得到函数 $\bar{G}(x)$ 的 B 微分中的元素

$$H(x) = (\max_{j\in J_{11}} f_{1j1}(x), \cdots, \max_{j\in J_{m1}} f_{mj1}(x), \cdots, \max_{j\in J_{1p}} f_{1jp}(x), \cdots, \max_{j\in J_{mp}} f_{mjp}(x))^{\mathrm{T}}.$$

如果

$$J(f_{1j_{11}1}(x), \cdots, f_{mj_{m1}1}(x), \cdots, f_{1j_{1p}p}(x), \cdots, f_{mj_{mp}p}(x))^{\mathrm{T}} \in \partial_B H(x),$$

那么

$$J(g_1(f_{1j_{11}1}(x), \cdots, f_{mj_{m1}1}(x)), \cdots, g_p(f_{1j_{1p}p}(x), \cdots, f_{mj_{mp}p}(x)))^{\mathrm{T}} \in \partial_B \bar{G}(x)$$

是 $\partial_B \bar{G}(x)$ 的一个元素.

本节的方法可应用于以下序补问题 (OCP) 的牛顿法, 序补问题是找出一个 $x^* \in \mathbf{R}^n$, 使得

$$F^j(x^*) \geqslant 0, \quad j = 1, \cdots, m,$$

$$\prod_{j=1}^{m} F_i^j(x^*) = 0, \quad i = 1, \cdots, n,$$

其中 $F^j(x), j=1,\cdots,m$ 是连续可微的, $F_i^j(x)$ 是 $F^j(x)$ 的第 i 个元素. 众所周知, 序补问题等价于求非光滑方程组

$$\max_{1\leqslant j\leqslant m}(-F^j(x))=0 \tag{4.5.20}$$

的解. 求解方程组 (4.5.20) 的牛顿法中, 在每异步迭代点处需要计算函数

$$\max_{1\leqslant j\leqslant m}(-F^j(x))$$

的 B 微分中一个元素.

第 5 章　拟可微函数及拟微分

拟可微函数是由 Demyanov 和 Rubinov 于 20 世纪 70 年代末引入的一类非光滑函数. 拟可微函数这一术语最早由 Pschenichny 提出, Demyanov 等借用了这一术语, Demyanov 的拟可微函数也是对 Pschenichny 相应拟可微函数的推广.

5.1　拟微分的定义及有关性质

定义 5.1.1　设 $f(x)$ 为 $\mathbf{R}^n$ 上的函数, 如果 $f(x)$ 是方向可微的, 且存在一对凸紧集 $\underline{\partial} f(x), \bar{\partial} f(x) \subset \mathbf{R}^n$, 使得其方向导数可表示为下述形式:

$$f'(x,d) = \max_{u\in\underline{\partial} f(x)} u^{\mathrm{T}} d + \min_{v\in\bar{\partial} f(x)} v^{\mathrm{T}} d, \quad \forall d \in \mathbf{R}^n, \tag{5.1.1}$$

则称 $f(x)$ 为拟可微函数. 集合对 $Df(x) = [\underline{\partial} f(x),\ \bar{\partial} f(x)]$ 称为 $f(x)$ 在 x 点处的拟微分, $\underline{\partial} f(x)$ 和 $\bar{\partial} f(x)$ 分别称为次微分和超微分. 如果 $\bar{\partial} f(x) = \{0\}$, 则称 $f(x)$ 是次可微的, 如果 $\underline{\partial} f(x) = \{0\}$, 则称 $f(x)$ 是超可微的.

从定义看出, 拟可微函数方向导数表示为两个凸紧集支撑函数的差, 事实上由式 (5.1.1) 易见, 方向导数可以表述为

$$\begin{aligned} f'(x,d) &= \max_{u\in\underline{\partial} f(x)} u^{\mathrm{T}} d - \max_{v\in-\bar{\partial} f(x)} v^{\mathrm{T}} d \\ &= \delta^*(d|\bar{\partial} f(x)) - \delta^*(d| - \bar{\partial} f(x)), \quad \forall d \in \mathbf{R}^n. \end{aligned}$$

显然, 凸函数 $f(x)$ 是拟可微函数, 它的拟微分为 $[\partial f(x), \{0\}]$, 此处 ∂ 代表凸函数的次微分, 因此凸函数也是次可微函数.

两个凸函数的差是拟可微函数. 设 $f(x) = g(x) - h(x)$, 其中 $g(x)$ 和 $h(x)$ 为 $\mathbf{R}^n$ 上凸函数, 易见, $f(x)$ 的方向导数可以表示为

$$\begin{aligned} f'(x,d) &= \max_{u\in\partial g(x)} u^{\mathrm{T}} d - \max_{v\in\partial h(x)} v^{\mathrm{T}} d \\ &= \max_{u\in\partial g(x)} u^{\mathrm{T}} d + \min_{v\in-\partial h(x)} v^{\mathrm{T}} d, \end{aligned}$$

于是 $f(x)$ 是拟可微函数, $[\partial g(x), -\partial h(x)]$ 为其拟微分, 其中 ∂ 代表凸函数的次微分.

为讨论拟微分的链锁法则, 我们规定 $\mathbf{R}^n$ 空间中集合对的加法和数乘运算如下:

$$[U_1, V_1] + [U_2, V_2] = [U_1 + U_2, V_1 + V_2],$$

$$\lambda[U, V] = \begin{cases} [\lambda U, \lambda V], & \lambda \geqslant 0, \\ [\lambda V, \lambda U], & \lambda < 0. \end{cases}$$

定理 5.1.1 (1) 设 $f_1(x)$ 和 $f_2(x)$ 为 $\mathbf{R}^n$ 上的拟可微函数, 则函数 $f_1(x)+f_2(x)$ 和 $f_1(x)f_2(x)$ 也是拟可微的, 且有

$$D(f_1(x) + f_2(x)) = Df_1(x) + Df_2(x), \tag{5.1.2}$$

$$D(f_1(x)f_2(x)) = f_1(x)Df_2(x) + f_2(x)Df_1(x), \tag{5.1.3}$$

即

$$\underline{\partial}(f_1(x) + f_2(x)) = \underline{\partial}f_1(x) + \underline{\partial}f_2(x),$$
$$\bar{\partial}(f_1(x) + f_2(x)) = \bar{\partial}f_1(x) + \bar{\partial}f_2(x),$$

$$\underline{\partial}(f_1(x)f_2(x)) = \begin{cases} f_1(x)\underline{\partial}f_2(x) + f_2(x)\underline{\partial}f_1(x), & f_1(x) \geqslant 0,\ f_2(x) \geqslant 0, \\ f_1(x)\bar{\partial}f_2(x) + f_2(x)\underline{\partial}f_1(x), & f_1(x) \leqslant 0,\ f_2(x) \geqslant 0, \\ f_1(x)\bar{\partial}f_2(x) + f_2(x)\bar{\partial}f_1(x), & f_1(x) \leqslant 0,\ f_2(x) \leqslant 0, \\ f_1(x)\underline{\partial}f_2(x) + f_2(x)\bar{\partial}f_1(x), & f_1(x) \geqslant 0,\ f_2(x) \leqslant 0, \end{cases}$$

$$\bar{\partial}(f_1(x)f_2(x)) = \begin{cases} f_1(x)\bar{\partial}f_2(x) + f_2(x)\bar{\partial}f_1(x), & f_1(x) \geqslant 0,\ f_2(x) \geqslant 0, \\ f_1(x)\underline{\partial}f_2(x) + f_2(x)\bar{\partial}f_1(x), & f_1(x) \leqslant 0,\ f_2(x) \geqslant 0, \\ f_1(x)\underline{\partial}f_2(x) + f_2(x)\underline{\partial}f_1(x), & f_1(x) \leqslant 0,\ f_2(x) \leqslant 0, \\ f_1(x)\bar{\partial}f_2(x) + f_2(x)\underline{\partial}f_1(x), & f_1(x) \geqslant 0,\ f_2(x) \leqslant 0. \end{cases}$$

(2) 设 $f(x)$ 是拟可微函数, 则对任意实数 λ, 函数 $\lambda f(x)$ 也是拟可微的, 且有

$$D(\lambda f(x)) = \lambda Df(x),$$

即

$$\underline{\partial}(\lambda f(x)) = \begin{cases} \lambda\underline{\partial}f(x), & \lambda \geqslant 0, \\ \lambda\bar{\partial}f(x), & \lambda < 0, \end{cases}$$

$$\bar{\partial}(\lambda f(x)) = \begin{cases} \lambda\bar{\partial}f(x), & \lambda \geqslant 0, \\ \lambda\underline{\partial}f(x), & \lambda < 0. \end{cases}$$

(3) 设 $f(x)$ 是拟可微函数, 且 $f(x) \neq 0$, 则函数 $\dfrac{1}{f(x)}$ 也是拟可微的, 其拟微分为

$$D\left(\frac{1}{f(x)}\right) = -\frac{1}{f^2(x)}Df(x),$$

即

$$\underline{\partial}\left(\frac{1}{f(x)}\right) = -\frac{1}{f^2(x)}\bar{\partial}f(x),$$

$$\bar{\partial}\left(\frac{1}{f(x)}\right) = -\frac{1}{f^2(x)}\underline{\partial}f(x).$$

证明　由方向导数的性质, 函数 $f_1(x)+f_2(x)$, $f_1(x)f_2(x)$, $\lambda f(x)$ 和 $\dfrac{1}{f(x)}$ 是方向可微的, 且方向导数为

$$(f_1+f_2)'(x,d) = f_1'(x;d) + f_2'(x;d), \tag{5.1.4}$$

$$(f_1\cdot f_2)'(x,d) = f_1(x)f_2'(x;d) + f_2(x)f_1'(x;d), \tag{5.1.5}$$

$$(\lambda f)'(x;d) = \lambda f'(x;d), \tag{5.1.6}$$

$$\left(\frac{1}{f}\right)'(x;d) = -\frac{1}{f^2(x)}f'(x;d). \tag{5.1.7}$$

根据拟可微函数定义, 利用拟微分表示方向导数, 代入式 (5.1.4)~(5.1.7) 得

$$\begin{aligned}(f_1+f_2)'(x;d) &= \max_{u\in\underline{\partial}f_1(x)} u^{\mathrm{T}}d + \min_{v\in\bar{\partial}f_1(x)} v^{\mathrm{T}}d + \max_{u\in\underline{\partial}f_2(x)} u^{\mathrm{T}}d + \min_{v\in\bar{\partial}f_2(x)} v^{\mathrm{T}}d \\ &= \max_{u\in\underline{\partial}f_1(x)+\underline{\partial}f_2(x)} u^{\mathrm{T}}d + \min_{v\in\bar{\partial}f_1(x)+\bar{\partial}f_2(x)} v^{\mathrm{T}}d,\end{aligned}$$

$$\begin{aligned}&(f_1\cdot f_2)'(x;d)\\ &= f_1(x)\left(\max_{u\in\underline{\partial}f_2(x)} u^{\mathrm{T}}d + \min_{v\in\bar{\partial}f_2(x)} v^{\mathrm{T}}\right) + f_2(x)\left(\max_{u\in\underline{\partial}f_1(x)} u^{\mathrm{T}}d + \min_{v\in\bar{\partial}f_1(x)} v^{\mathrm{T}}d\right)\\ &= \begin{cases} \max\limits_{u\in f_1(x)\underline{\partial}f_2(x)+f_2(x)\underline{\partial}f_1(x)} u^{\mathrm{T}}d + \min\limits_{v\in f_1(x)\bar{\partial}f_2(x)+f_2(x)\bar{\partial}f_1(x)} v^{\mathrm{T}}d, & f_1(x)\geqslant 0, f_2(x)\geqslant 0,\\ \max\limits_{u\in f_1(x)\bar{\partial}f_2(x)+f_2(x)\underline{\partial}f_1(x)} u^{\mathrm{T}}d + \min\limits_{v\in f_1(x)\underline{\partial}f_2(x)+f_2(x)\bar{\partial}f_1(x)} v^{\mathrm{T}}d, & f_1(x)\leqslant 0, f_2(x)\geqslant 0,\\ \max\limits_{u\in f_1(x)\bar{\partial}f_2(x)+f_2(x)\bar{\partial}f_1(x)} u^{\mathrm{T}}d + \min\limits_{v\in f_1(x)\underline{\partial}f_2(x)+f_2(x)\bar{\partial}f_1(x)} v^{\mathrm{T}}d, & f_1(x)\leqslant 0, f_2(x)\leqslant 0,\\ \max\limits_{u\in f_1(x)\underline{\partial}f_2(x)+f_2(x)\bar{\partial}f_1(x)} u^{\mathrm{T}}d + \min\limits_{v\in f_1(x)\bar{\partial}f_2(x)+f_2(x)\underline{\partial}f_1(x)} v^{\mathrm{T}}d, & f_1(x)\geqslant 0, f_2(x)\leqslant 0,\end{cases}\end{aligned}$$

$$\begin{aligned}(\lambda f)'(x;d) &= \lambda\left(\max_{u\in\underline{\partial}f(x)} u^{\mathrm{T}}d + \min_{v\in\bar{\partial}f(x)} v^{\mathrm{T}}d\right)\\ &= \begin{cases}\max\limits_{u\in\lambda\underline{\partial}f(x)} u^{\mathrm{T}}d + \min\limits_{v\in\lambda\bar{\partial}f(x)} v^{\mathrm{T}}d, & \lambda\geqslant 0,\\ \max\limits_{u\in\lambda\bar{\partial}f(x)} u^{\mathrm{T}}d + \min\limits_{v\in\lambda\underline{\partial}f(x)} v^{\mathrm{T}}d, & \lambda<0,\end{cases}\end{aligned}$$

$$\begin{aligned}\left(\frac{1}{f}\right)'(x;d) &= -\frac{1}{f^2(x)}\left(\max_{u\in\underline{\partial}f(x)} u^{\mathrm{T}}d + \min_{v\in\bar{\partial}f(x)} v^{\mathrm{T}}d\right)\\ &= \max_{u\in-\frac{1}{f^2(x)}\bar{\partial}f(x)} u^{\mathrm{T}}d + \min_{v\in-\frac{1}{f^2(x)}\underline{\partial}f(x)} v^{\mathrm{T}}d.\end{aligned}$$

根据拟微分定义和式 (5.1.4)~(5.1.7), 得式 (5.1.2) 和式 (5.1.3). 定理得证.

利用定理 5.1.1 还可以得到下面的拟微分运算法则:

$$D(f_1(x)-f_2(x))=Df_1(x)-Df_2(x),$$

$$D\left(\frac{f_1(x)}{f_2(x)}\right)=\frac{1}{f_2^2(x)}(f_2(x)Df_1(x)-f_1(x)Df_2(x)).$$

定理 5.1.2 设 $f_i(x), i=1,\cdots,m$ 为 $\mathbf{R}^n$ 上的方向可微函数, 则函数 $f(x)=\max\limits_{1\leqslant i\leqslant m} f_i(x)$ 也是方向可微的, 且方向导数为

$$f'(x;d)=\max_{i\in I(x)} f_i'(x;d), \quad d\in \mathbf{R}^n, \tag{5.1.8}$$

其中

$$I(x)=\{1\leqslant i\leqslant m | f_i(x)=f(x)\}.$$

证明 对固定的 $d\in\mathbf{R}^n$, 根据 $f_i(x)$ 的方向可微性有

$$f_i(x+td)=f_i(x)+tf_i'(x;d)+o_i(t), \tag{5.1.9}$$

当 t 充分小时

$$f_i(x+td)<f(x+td), \quad \forall i\in\{1,\cdots,m\}\backslash I(x),$$

故

$$f(x+td)=\max_{1\leqslant i\leqslant m} f_i(x+td)=\max_{i\in I(x)} f_i(x;d). \tag{5.1.10}$$

结合式 (5.1.9) 和式 (5.1.10) 得

$$f(x+td)-f(x)=\max_{i\in I(x)}(tf_i'(x;d)+o_i(t)),$$

于是

$$f(x+td)-f(x)-t\max_{i\in I(x)} f_i'(x;d)=o(t),$$

上式两边除以 t, 并令 $t\to 0$, 即得式 (5.1.8). 定理得证.

定理 5.1.3 设 $f_i(x), i=1,\cdots,m$ 为 $\mathbf{R}^n$ 上的拟可微函数, 则它们的极大值函数 $f(x)=\max\limits_{1\leqslant i\leqslant m} f_i(x)$ 也是拟可微的, 其拟微分 $[\underline{\partial} f(x), \bar{\partial} f(x)]$ 为

$$\underline{\partial} f(x)=\operatorname{co}\bigcup_{k\in I(x)}\left(\underline{\partial} f_k(x)-\sum_{i\in I(x)\backslash\{k\}}\bar{\partial} f_i(x)\right),$$

$$\bar{\partial} f(x)=\sum_{i\in I(x)}\bar{\partial} f_i(x),$$

其中

$$I(x)=\{1\leqslant i\leqslant m|f_i(x)=f(x)\}.$$

证明　$f_i(x)$ 是拟可微的, 简记其拟微分 $[\underline{\partial}f_i(x),\bar{\partial}f_i(x)]$ 为 $\underline{\partial}f_i(x)=A_i,\bar{\partial}f_i(x)=B_i$. 不失一般性, 假设 $I(x)=\{1,\cdots,m\}$. 根据定理 5.1.2, $f(x)$ 是方向可微的, 其方向导数为

$$f'(x,d)=\max_{1\leqslant i\leqslant m}(\max_{u\in A_i}u^{\mathrm{T}}d+\min_{v\in B_i}v^{\mathrm{T}}d),\quad d\in\mathbf{R}^n. \tag{5.1.11}$$

首先讨论 $m=2$ 情形. 设 a,b,c,d,r 是实数, 因为

$$\max\{a+b,c+d\}=\max\{a+b-r,c+d-r\}+r,$$

取 $r=b+d$, 则有

$$\max\{a+b,c+d\}=\max\{a-d,c-b\}+(b+d).$$

于是, 由式 (5.1.11) 得

$$\begin{aligned}H_2=&\max_{i=1,2}(\max_{u\in A_i}u^{\mathrm{T}}d+\min_{v\in B_i}v^{\mathrm{T}}d)\\=&\max\left\{\max_{u\in A_1}u^{\mathrm{T}}d-\min_{v\in B_2}v^{\mathrm{T}}d,\max_{u\in A_2}u^{\mathrm{T}}d-\min_{v\in B_1}v^{\mathrm{T}}d\right\}\\&+\min_{u\in B_1}u^{\mathrm{T}}d+\min_{v\in B_2}v^{\mathrm{T}}d.\end{aligned}\tag{5.1.12}$$

式 (5.1.12) 可变形为

$$\begin{aligned}H_2=&\max\left\{\max_{u\in A_1}u^{\mathrm{T}}d+\max_{v\in -B_2}v^{\mathrm{T}}d,\ \max_{u\in A_2}u^{\mathrm{T}}d+\max_{v\in -B_1}v^{\mathrm{T}}d\right\}\\&+\min_{u\in B_1}u^{\mathrm{T}}d+\min_{v\in B_2}v^{\mathrm{T}}d\\=&\max\left\{\max_{u\in A_1-B_2}u^{\mathrm{T}}d,\ \max_{u\in A_2-B_1}u^{\mathrm{T}}d\right\}+\min_{v\in B_1+B_2}v^{\mathrm{T}}d,\end{aligned}$$

注意到

$$\max\left\{\max_{u\in A}u^{\mathrm{T}}d,\ \max_{u\in B}u^{\mathrm{T}}d\right\}=\max_{u\in\mathrm{co}\{A\cup B\}}u^{\mathrm{T}}d,$$

于是

$$H_2=\max_{u\in\mathrm{co}\{(A_1-B_2)\cup(A_2-B_1)\}}u^{\mathrm{T}}d+\min_{v\in B_1+B_2}v^{\mathrm{T}}d.$$

当 $m=l$ 时, 假设定理结论成立, 即

$$H_l=\max_{1\leqslant i\leqslant l}\left(\max_{u\in A_i}u^{\mathrm{T}}d+\min_{v\in B_i}v^{\mathrm{T}}d\right)=\max_{u\in U}u^{\mathrm{T}}d+\min_{v\in V}v^{\mathrm{T}}d,$$

其中

$$U = \text{co}\left\{A_k - \sum_{\substack{1\leqslant i\leqslant l\\ i\neq k}} B_i \,|1 \leqslant i \leqslant l\right\},$$

$$V = \sum_{i=1}^{l} B_i,$$

于是有

$$\begin{aligned} H_{l+1} &= \max_{1\leqslant i\leqslant l+1} \left(\max_{u\in A_i} u^{\mathrm{T}}d + \min_{v\in B_i} v^{\mathrm{T}}d\right) \\ &= \max\left\{H_l,\ \max_{u\in A_{l+1}} u^{\mathrm{T}}d + \min_{v\in B_{l+1}} v^{\mathrm{T}}d\right\} \\ &= \max\left\{\max_{u\in U} u^{\mathrm{T}}d + \min_{v\in V} v^{\mathrm{T}}d + \max_{u\in A_{l+1}} u^{\mathrm{T}}d + \min_{v\in B_{l+1}} v^{\mathrm{T}}d\right\}. \end{aligned} \tag{5.1.13}$$

记 $H_{l+1} = \max\limits_{u\in A} u^{\mathrm{T}}d + \min\limits_{v\in B} v^{\mathrm{T}}d$, 利用前面的方法, 由式 (5.1.13) 得

$$\begin{aligned} A &= \text{co}\,\{U - B_{l+1},\ A_{l+1} - V\} \\ &= \text{co}\left\{A_k - \sum_{\substack{1\leqslant i\leqslant l+1\\ i\neq k}} B_i | k = 1,\cdots,l+1\right\}, \\ B &= V + B_{l+1} = \sum_{i=1}^{l+1} B_i. \end{aligned}$$

定理得证.

设 $f_i(x)$ 为拟可微函数, 则

$$f(x) = \min_{1\leqslant i\leqslant m} f_i(x)$$

也是拟可微的, 其拟微分 $[\underline{\partial} f(x),\ \bar{\partial} f(x)]$ 为

$$\underline{\partial} f(x) = \sum_{i\in I(x)} \underline{\partial} f_i(x),$$

$$\bar{\partial} f(x) = \text{co}\left\{\bar{\partial} f_k(x) - \sum_{i\in I(x)\setminus\{k\}} \underline{\partial} f_i(x) | k \in I(x)\right\},$$

其中

$$I(x) = \{1 \leqslant i \leqslant m | f_i(x) = f(x)\}.$$

注意到

$$\min_{1\leqslant i\leqslant m} f_i(x) = -\max_{1\leqslant i\leqslant m} (-f_i(x)),$$

由定理 5.1.3 即可得到上述公式.

定理 5.1.4　设 $f(y)$ 为 $\mathbf{R}^m$ 上的拟可微函数, 且一致方向可微, $h_i(x), i=1,\cdots,m$ 为 $\mathbf{R}^n$ 上的拟可微函数, 则复合函数

$$F(x)=f(h_1(x),\cdots,h_m(x))$$

是拟可微的, 其拟微分 $\left[\underline{\partial}F(x),\ \bar{\partial}F(x)\right]$ 为

$$\underline{\partial}F(x)=\left\{p\middle|p=\sum_{i=1}^{m}(u^{(i)}(\lambda_i+\mu_i)-\underline{u}^{(i)}\lambda_i-\bar{u}^{(i)}\mu_i),u=(u^{(1)},\cdots,u^{(m)})\in\underline{\partial}f(y),\right.$$
$$\left.\lambda_i\in\underline{\partial}h_i(x),\mu_i\in\bar{\partial}h_i(x)\right\},$$

$$\bar{\partial}F(x)=\left\{l\middle|l=\sum_{i=1}^{m}(u^{(i)}(\lambda_i+\mu_i)+\underline{u}^{(i)}\lambda_i+\bar{u}^{(i)}\mu_i),u=(u^{(1)},\cdots,u^{(m)})\in\bar{\partial}f(y),\right.$$
$$\left.\lambda_i\in\underline{\partial}h_i(x),\mu_i\in\bar{\partial}h_i(x)\right\},$$

其中, $\underline{u}=(\underline{u}^{(1)},\cdots,\underline{u}^{(m)})$ 和 $\bar{u}=(\bar{u}^{(1)},\cdots,\bar{u}^{(m)})$ 是满足下式

$$\underline{u}\leqslant u\leqslant\bar{u},\ \forall u\in\underline{\partial}f(y)\bigcup(-\bar{\partial}f(y))$$

的任意向量.

定理证明可参考 Demyanov 和 Rubinov 的专著, 这里从略.

以上定理说明, 拟可微性质关于四则运算、极大 (小) 值运算以及复合运算是封闭的.

5.2　拟可微函数类及有关性质

拟可微函数是一类非常广的非光滑函数, 其中极大值函数及其复合是最适合利用拟微分工具来处理的.

5.2.1　极大值复合函数

设

$$f(x)=g\big(\max_{j\in J_1}f_{1j}(x),\cdots,\max_{j\in J_m}f_{mj}(x)\big),\quad x\in\mathbf{R}^n,\tag{5.2.1}$$

其中 $g(y_1,\cdots,y_m)$ 和 $f_{ij}(x)$ 分别为 $\mathbf{R}^m$ 和 $\mathbf{R}^n$ 上的连续可微函数, $J_i,i=1,\cdots,m$ 为有限指标集. 由定理 5.1.3 和定理 5.1.4 可知 $f(x)$ 为拟可微函数, 以下利用定义直接推导 $f(x)$ 的拟微分. 记

$$y_i(x)=\max_{j\in J_i}f_{ij}(x),\quad x\in\mathbf{R}^n,\quad i=1,\cdots,m,$$

对于 $d\in\mathbf{R}^n, t\geqslant 0$, 有

$$y_i(x+td)=y_i(x)+ty_i'(x;d)+o_i(t,d),$$

$$y_i'(x;d)=\max_{j\in J_i(x)}\nabla f_{ij}(x)^{\mathrm{T}}d,\quad i=1,\cdots,m,$$

其中

$$J_i(x)=\{j\in I_i|f_{ij}(x)=y_i(x)\}.$$

由 $o_i(t,d)/t\to 0$ 得

$$f(x+td)=f(x)+t\left(\sum_{i=1}^{m}\frac{\partial F(y_1,\cdots,y_m)}{\partial y_i}y_i'(x;d)\right)+o(t,d). \tag{5.2.2}$$

记

$$I_+(x)=\left\{1\leqslant i\leqslant m\left|\frac{\partial F(y_1,\cdots,y_m)}{\partial y_i}\geqslant 0\right.\right\},$$

$$I_-(x)=\left\{1\leqslant i\leqslant m\left|\frac{\partial F(y_1,\cdots,y_m)}{\partial y_i}<0\right.\right\},$$

由式 (5.2.2) 得

$$f(x+td)=f(x)+t\left(\sum_{i\in I_+(x)}\max_{j\in J_i(x)}\frac{\partial F(y_1,\cdots,y_m)}{\partial y_i}\nabla f_{ij}(x)^{\mathrm{T}}d\right.$$
$$\left.+\sum_{i\in I_-(x)}\min_{j\in J_i(x)}\frac{\partial F(y_1,\cdots,y_m)}{\partial y_i}\nabla f_{ij}(x)^{\mathrm{T}}d\right)+o(t,d),$$

于是

$$f'(x;d)=\sum_{i\in I_+(x)}\max_{j\in J_i(x)}\frac{\partial F(y_1,\cdots,y_m)}{\partial y_i}\nabla f_{ij}(x)^{\mathrm{T}}d$$
$$+\sum_{i\in I_-(x)}\min_{j\in J_i(x)}\frac{\partial F(y_1,\cdots,y_m)}{\partial y_i}\nabla f_{ij}(x)^{\mathrm{T}}d.$$

根据拟微分的定义, $f(x)$ 的拟微分 $[\underline{\partial}f(x),\ \bar{\partial}f(x)]$ 为

$$\underline{\partial}f(x)=\mathrm{co}A(x),\quad \bar{\partial}f(x)=\mathrm{co}B(x),$$

其中

$$A(x)=\left\{u\in\mathbf{R}^n\left|u=\sum_{i\in I_+(x)}\frac{\partial F(y_1,\cdots,y_m)}{\partial y_i}\nabla f_{ij}(x)^{\mathrm{T}}d, j\in I_i(x)\right.\right\},$$

$$B(x)=\left\{v\in\mathbf{R}^n\left|v=\sum_{i\in I_-(x)}\frac{\partial F(y_1,\cdots,y_m)}{\partial y_i}\nabla f_{ij}(x)^{\mathrm{T}}d, j\in I_i(x)\right.\right\}.$$

根据上面例子可以看出, 函数

$$f(x)=\max_{i\in I}f_i(x)+\min_{j\in J}g_j(x)$$

是拟可微的, 其中 $f_i(x), i\in I, g_i(x), j\in J$ 为连续可微函数, I, J 为有限指标集, 其拟微分 $[\underline{\partial}f(x),\bar{\partial}f(x)]$ 为

$$\underline{\partial}f(x)=\mathrm{co}\left\{\nabla f_i(x)\,|i\in I(x)\right\},$$
$$\bar{\partial}f(x)=\mathrm{co}\left\{\nabla g_j(x)\,|j\in J(x)\right\},$$

其中

$$I(x)=\left\{i\in I\,\middle|\,f_i(x)=\max_{i\in I}f_i(x)\right\},$$
$$J(x)=\left\{i\in J\,\middle|\,g_j(x)=\min_{j\in J}g_j(x)\right\}.$$

5.2.2 一维函数

命题 5.2.1　设 $f(x)$ 为 $\mathbf{R}$ 上的函数, $f(x)$ 在点 x 是拟可微的充要条件是 $f(x)$ 在点 x 是方向可微的.

证明　只需考虑充分性. 设 $f(x)$ 在点 x 是方向可微的, 令

$$\begin{aligned}a_1&=\min\{f'(x;1),-f'(x;-1)\},\\a_2&=\max\{f'(x;1),-f'(x;-1)\},\\b&=\max\{0,-f'(x;1)-f'(x;-1)\},\\A&=[a_1,\ a_2],\quad B=[-b,b].\end{aligned}$$

利用定义可以验证 $[A,\ B]$ 为 $f(x)$ 在点 x 的拟微分, 事实上

$$\max_{u\in A}ud+\min_{v\in B}vd=a_2d-bd=f'(x;d),\quad d\geqslant 0,$$

$$\max_{u\in A}ud+\min_{v\in B}vd=a_1d+bd=f'(x;d),\quad d<0.$$

命题得证.

根据命题 5.2.1, 我们给出一个拟可微函数而非局部 Lipschitz 函数的例子. 设

$$f(x)=\begin{cases}x^{\frac{3}{2}}\sin\dfrac{1}{x}, & x>0,\\ x, & x\leqslant 0.\end{cases}$$

可以验证, $f(x)$ 在 $\mathbf{R}$ 上是方向可微的, 根据命题 5.2.1, 它是拟可微的, 但 $f(x)$ 在 $x=0$ 附近不是 Lipschitz 的. 事实上, 拟可微函数与局部 Lipschitz 函数是两类互不包含的不可微函数.

5.2.3 极值条件

定理 5.2.1 设 $f(x)$ 为 $\mathbf{R}^n$ 上的拟可微函数, 如果 $f(x)$ 在 $x^* \in \mathbf{R}^n$ 处达到极小值, 则有

$$-\bar{\partial} f(x^*) \subset \underline{\partial} f(x^*). \tag{5.2.3}$$

证明 因为 $f(x)$ 是方向可微的, x^* 是极小值点, 则有

$$f'(x;d) \geqslant 0, \quad \forall d \in \mathbf{R}^n, \tag{5.2.4}$$

若不然, 则存在 $d_1 \in \mathbf{R}^n$, 使得 $f'(x^*;d_1) < 0$, 根据方向导数定义, 当 $t > 0$ 充分小时

$$f(x^* + td_1) < f(x^*),$$

这与 x^* 是极小值点矛盾. 由式 (5.2.4) 及拟微分的定义可得

$$0 \leqslant \max_{u \in \underline{\partial} f(x^*)} u^{\mathrm{T}} d + \min_{v \in \bar{\partial} f(x^*)} v^{\mathrm{T}} d, \quad \forall d \in \mathbf{R}^n,$$

于是

$$\max_{v \in -\bar{\partial} f(x^*)} v^{\mathrm{T}} d \leqslant \max_{u \in \underline{\partial} f(x^*)} u^{\mathrm{T}} d, \quad \forall d \in \mathbf{R}^n, \tag{5.2.5}$$

根据凸紧集与支撑函数的关系和式 (5.2.5) 得式 (5.2.3). 定理得证.

定理 5.2.2 设 $f(x)$ 是 $\mathbf{R}^n$ 上的拟可微函数, 如果 $f(x)$ 在 $x^* \in \mathbf{R}^n$ 处达到极大值, 则有

$$-\underline{\partial} f(x^*) \subset \bar{\partial} f(x^*). \tag{5.2.6}$$

证明 $f(x)$ 在 x^* 处达到极大值, 类似定理 5.2.1 的证明, 可得

$$f'(x;d) \leqslant 0, \quad \forall d \in \mathbf{R}^n,$$

类似地, 也可证明式 (5.2.6) 成立. 定理得证.

与 Clarke 广义梯度形式的极值条件不同, 拟微分形式的极大值与极小值必要条件在形式上是有所不同的, 通常将满足式 (5.2.3) 和式 (5.2.6) 的点 x^* 分别称为极小驻点和极大驻点.

定理 5.2.3 设 $f(x)$ 是 $\mathbf{R}^n$ 上拟可微和一致方向可微的, 如果

$$-\bar{\partial} f(x^*) \subset \mathrm{int}\underline{\partial} f(x^*),$$

则 x^* 为 $f(x)$ 的严格极小点.

证明 由

$$-\bar{\partial} f(x^*) \subset \mathrm{int}\underline{\partial} f(x^*)$$

可知, 存在 $r>0$, 使得

$$B(0,r)\subset \underline{\partial} f(x^*)+v,\ \forall v\in\bar{\partial} f(x^*),$$

于是有

$$\max_{u\in\underline{\partial} f(x^*)+v} u^{\mathrm{T}}d\geqslant r,\quad \forall v\in\bar{\partial} f(x^*),\quad \|d\|=1.$$

关于 $v\in\bar{\partial} f(x^*)$ 取极小, 有

$$\min_{v\in\bar{\partial} f(x^*)}\max_{u\in\underline{\partial} f(x^*)+v} u^{\mathrm{T}}d\geqslant r,\quad \|d\|=1. \tag{5.2.7}$$

对式 (5.2.7) 变形得

$$\begin{aligned}\min_{v\in\bar{\partial} f(x^*)}\max_{u\in\underline{\partial} f(x^*)+v} u^{\mathrm{T}}d&=\min_{v\in\bar{\partial} f(x^*)}\Big(\max_{u\in\underline{\partial} f(x^*)} u^{\mathrm{T}}d+v^{\mathrm{T}}d\Big)\\&=\max_{u\in\underline{\partial} f(x^*)} u^{\mathrm{T}}d+\min_{v\in\bar{\partial} f(x^*)} v^{\mathrm{T}}d\\&=f'(x^*;d)\geqslant r,\quad \|d\|=1.\end{aligned}$$

再由 $f(x)$ 的一致方向可微性及上式, 则存在 $t_0>0$, 使得

$$f(x^*+td)\geqslant f(x^*)+\frac{1}{2}tr,\quad \forall t\in[0,t_0],$$

这说明 x^* 是 $f(x)$ 的严格极小值点. 定理得证.

5.3　凸紧集的差

拟微分定义与方向导数有直接联系, 因此, 拟微分是目前各种非凸次微分中最容易计算的一种. 本节将引入两个凸紧集的差, 除具有拟可微分析自身意义外, 还可用于计算 Clarke 广义梯度.

5.3.1　定义

设 U 和 V 是 $\mathbf{R}^n$ 上的凸紧集, $T\subset\mathbf{R}^n$, 如果 $\mathbf{R}^n\backslash T$ 是一个零测度集, 则称 T 为关于 $\mathbf{R}^n$ 的满测度集.

给定满测度集 T, 使得在每一点 $x\in T$ 处 $\delta^*(x|U)$ 和 $\delta^*(x|V)$ 可微, 定义 U 和 V 的差 $U\dot{-}V$ 如下:

$$U\dot{-}V=\mathrm{clco}\{\nabla\delta^*(x|U)-\nabla\delta^*(x|V)|x\in T\}. \tag{5.3.1}$$

$S\subset\mathbf{R}^n$ 在点 $x\in\mathbf{R}^n$ 的最大面 $G_x(S)$ 和最小面 $\tilde{G}_x(S)$ 定义如下:

$$G_x(S)=\{s\in S|s^{\mathrm{T}}x=\max_{s\in S} s^{\mathrm{T}}x\},$$

$$\tilde{G}_x(S) = \{s \in S | s^{\mathrm{T}} x = \min_{s \in S} s^{\mathrm{T}} x\},$$

$G_x(S)$ 和 $\tilde{G}_x(S)$ 中的元素分别用 $g_x(S)$ 和 $\tilde{g}_x(S)$ 表示. 注意到 $x \in T$, 可以推出 $G_x(U)$ 和 $\tilde{G}_x(V)$ 都是单点集, 因此 $U \dot{-} V$ 可以等价地表示为

$$U \dot{-} V = \mathrm{clco}\{g_x(U) - \tilde{g}_x(V) | x \in T\}.$$

首先, 我们要说明按式 (5.3.1) 定义的合理性, 即 $U \dot{-} V$ 不依赖于集合 T 的选择. 假设 T_{UV} 是所有 $\nabla\delta^*(x|U)$ 和 $\nabla\delta^*(x|V)$ 存在的点集, $T \subset T_{UV}$ 是一个满测度集, 于是

$$T_{UV} \subset \mathbf{R}^n = \mathrm{cl}T.$$

记

$$\begin{aligned} W_1 &= \mathrm{cl}\{\nabla\delta^*(x|U) - \nabla\delta^*(x|V) | x \in T_{UV}\}, \\ W_2 &= \mathrm{cl}\{\nabla\delta^*(x|U) - \nabla\delta^*(x|V) | x \in T\}. \end{aligned}$$

显然, $W_2 \subset W_1$. 另一方面, 设 $x \in T_{UV}$, 则有 $T_{UV} \subset \mathrm{cl}T$, 于是存在序列 $\{x_k\}$ 满足 $x_k \in T$, $x_k \to x$. 既然梯度 $\nabla\delta^*(x|U)$ 存在, 那么

$$\nabla\delta^*(x_k|U) \to \nabla\delta^*(x|U).$$

同理

$$\nabla\delta^*(x_k|V) \to \nabla\delta^*(x|V).$$

因此

$$\nabla\delta^*(x|U) - \nabla\delta^*(x|V) \in W_2, \quad W_1 \subset W_2, \quad W_1 = W_2.$$

这样, 如果

$$T_1, T_2 \subset T_{UV},$$

T_1 和 T_2 是满测度的, 有

$$\mathrm{clco}\{\nabla\delta^*(x|U) - \nabla\delta^*(x|V) | x \in T_1\} = \mathrm{clco}\{\nabla\delta^*(x|U) - \nabla\delta^*(x|V) | x \in T_2\}.$$

因此, 式 (5.3.1) 不依赖于 T 的选择, 即 $U \dot{-} V$ 是适定的. $U \dot{-} V$ 称为集合 U 和 V 的 Demyanov 差.

设 U 和 V 是 $\mathbf{R}^n$ 上的凸紧集, 集合的差 $U \ddot{-} V$ 定义为

$$U \ddot{-} V = \mathrm{co} \bigcup_{x \neq 0} (G_x(U) - G_x(V)).$$

容易验证集合 $U \ddot{-} V$ 是闭的. 称为集合 U 和 V 的 Rubinov 差.

5.3.2 拟微分表示 Clarke 广义梯度

下面基于集合的差, 利用拟微分表示 B 微分和 Clarke 广义梯度.

设 $f(x)$ 是 $\mathbf{R}^n$ 上的拟可微函数, 如果下述条件成立:

(i) 集合 $\mathbf{R}^n\backslash T$ 的测度为零;

(ii) 对于任意的 $g\in T$, 集合 $G_g(\underline{\partial}f(x))$ 和 $\tilde{G}_g(\overline{\partial}f(x))$ 是单点集.

则称集合 $T\subset\mathbf{R}^n$ 为关于拟微分 $[\underline{\partial}f(x),\overline{\partial}f(x)]$ 满足性质 (ε).

设 $X\subset\mathbf{R}^n$ 为开集, $x\in X$, 定义 X 到 $\mathbf{R}^m$ 的一族函数类 $M(x)$ 如下: 如果 $f(x)\in M(x)$, 则满足以下条件:

(a) $f(x)$ 在 x 的邻域 $B(x,\delta)$ 是局部 Lipschitz 的, 这说明 D_f 关于 $B(x,\delta)$ 是满测度的, 即 $B(x,\delta)\backslash D_f$ 的测度为零;

(b) $f(x)$ 在 x 是拟可微的;

(c) 存在关于 $B(x,\delta)$ 是满测度的 $Q\subset D_f$ 和关于拟微分 $[\underline{\partial}f(x),\overline{\partial}f(x)]$ 满足 (ε) 性质的 T, 且由条件

$$g_k\to g,\quad t_k\to 0^+,\quad x_k=x+t_kg_k\in Q,\quad g\in T,$$

可以得到

$$\nabla f(x_k)\to g_g(\underline{\partial}f(x))+\tilde{g}_g(\overline{\partial}f(x)).$$

下面的定理给出了 Clarke 广义梯度和拟微分之间的关系.

定理 5.3.1 给定一点 $x\in\mathbf{R}^n$, 如果 $f(x)$ 属于 $\mathbf{R}^n$ 上的函数类 $M(x)$, 则有

$$\underline{\partial}f(x)\dot{-}(-\overline{\partial}f(x))\subset\partial f(x). \tag{5.3.2}$$

证明 取集合 $Q\subset B(x,\delta)\bigcap D_f$ 满足

$$\mathrm{meas}(B(x,\delta)\backslash Q)=0, \tag{5.3.3}$$

其中 meas 代表 Lebesgue 测度, 另取 $T\subset\mathbf{R}^n$ 满足 $\mathrm{meas}(\mathbf{R}^n\backslash T)=0$ 及具有性质 (ε). 记

$$\partial_Tf(x)=\mathrm{clco}\{\xi|x_k=x+t_kg_k\in Q,g_k\to g\in T,t_k\to 0,\nabla f(x_k)\to\xi\}.$$

设 $v\in\partial_TF(x)$, 那么存在序列 $\{g_k\}$ 和 $\{t_k\}$ 满足

$$g_k\to g,\quad t_k\to 0^+,\quad x_k=x+t_kg_k\in Q,\quad g\in T,\quad \nabla F(x_k)\to v.$$

另一方面, 由 $F(x)\in M(x)$ 可得

$$\nabla F(x_k)\to g_g(\underline{\partial}F(x))+\tilde{g}_g(\overline{\partial}F(x)),$$

即

$$v = g_g(\underline{\partial}F(x)) + \tilde{g}_g(\overline{\partial}F(x)).$$

这样, 有

$$\partial_T F(x) \subset \text{clco}\{\, g_g(\underline{\partial}F(x)) + \tilde{g}_g(\overline{\partial}F(x)) \,|\, g \in T\} = \underline{\partial}F(x) \dot{-} (-\overline{\partial}F(x)). \tag{5.3.4}$$

反过来, 假设 $g \in \mathbf{R}^n$, $t_k \to 0^+$, $\varepsilon_k \to 0^+$, 由式 (5.3.3) 可得

$$Q \bigcap (x + t_k B(g, \varepsilon_k)) \neq \varnothing,$$

那么存在 $g_k \in B(g, \varepsilon_k)$ 满足

$$x + t_k g_k \in Q. \tag{5.3.5}$$

因此, 能找到一个序列 $\{g_k\}$ 满足 $g_k \to g$, 使得式 (5.3.5) 成立. 取 $g \in T$, 考虑元素

$$v = g_g(\underline{\partial}f(x)) + \tilde{g}_g(\overline{\partial}f(x)),$$

序列 $\{g_k\}$ 和 $\{t_k\}$ 满足 $g_k \to g$, $t_k \to 0^+$, 式 (5.3.5) 成立. 根据 $M(x)$ 的定义, $\nabla f(x) \to v$ 是成立的, 即

$$v \in \text{cl}\{\xi | x_k = x + t_k g_k \in Q, g_k \to g \in T, t_k \to 0^+, \nabla f(x_k) \to \xi\} = \partial_T f(x).$$

由此得

$$\underline{\partial}f(x) \dot{-} (-\overline{\partial}f(x)) = \partial_T f(x). \tag{5.3.6}$$

根据式 (5.3.4) 和式 (5.3.6) 得

$$\underline{\partial}f(x) \dot{-} (-\overline{\partial}f(x)) = \partial_T f(x).$$

注意到 $\partial f(x)$ 是闭集, 有 $\partial_T f(x) \subset \partial f(x)$, 式 (5.3.2) 成立. 定理得证.

定理 5.3.2 设 $U, V \subset \mathbf{R}^n$, 则有

$$U \dot{-} V = \partial(\delta^*(x|U) - \delta^*(x|V))|_{x=0}. \tag{5.3.7}$$

定理证明可参考 Demyanov 和 Rubinov 的专著, 这里从略. 尽管拟微分不是唯一的, 但式 (5.3.8) 说明次微分与负超微分的差是唯一确定的.

除具有 $M(x)$ 的条件 (a), (b) 外, 函数族 $\tilde{M}(x)$ 还要求满足下述条件:

(d) 存在 $B(x, \delta)$ 上的满测度的集合 $Q \subset D_F$ 和拟微分 $[\underline{\partial}f(x), \overline{\partial}f(x)]$, 使得由关系式

$$g_k \to g, \quad t_k \downarrow 0^+, \quad x_k = x + t_k g_k \in D_F, \quad \nabla f(x_k) \to v$$

可以得到

$$v \in G_g(\underline{\partial} f(x)) + \tilde{G}_g(\overline{\partial} f(x)).$$

定理 5.3.3　设 $\mathbf{R}^n$ 上的函数 $f(x)$ 包含在函数族 $\tilde{M}(x)$ 中, $[\underline{\partial} f(x), \overline{\partial} f(x)]$ 是满足条件 (d) 中的一个拟微分, 那么下述关系成立:

$$\partial f(x) \subset \underline{\partial} f(x) \ddot{-} (-\overline{\partial} f(x)). \tag{5.3.8}$$

证明　设

$$\xi \in \partial f(x) = \{v | x_k \in D_f, x_k \to x, \nabla f(x_k) \to v\},$$

序列 $\{v_k\}$ 满足

$$x_k \in D_f, \quad x_k \to x, \quad \nabla f(x_k) \to \xi.$$

记

$$\alpha_k = \|x_k - x\|, \quad g_k = (x_k - x)/\alpha_k.$$

不失一般性, 假设

$$\alpha_k \downarrow 0, \quad \lim_{k\to\infty} g_k = g,$$

根据性质 (d) 有

$$\xi \in G_g(\underline{\partial} f(x) + \tilde{G}_g(\overline{\partial} f(x)) = G_g(\underline{\partial} f(x)) - G_g(-\overline{\partial} g(x)),$$

由此得

$$\partial f(x) \subset \underline{\partial} f(x) \ddot{-} (-\overline{\partial} f(x)).$$

定理得证.

5.3.3 多面体公式

以下给出多面体情形下集合差的计算.

定理 5.3.4　设

$$U = \mathrm{co}\{u_i | i \in I\}, \quad V = \mathrm{co}\{v_j | j \in J\},$$

其中 $u_i, v_j \in \mathbf{R}^n$, I, J 为有限指标集. 给定一对指标 $i \in I$, $j \in J$, 定义线性不等式系统:

$$\begin{aligned} (L_{ij}) \quad & (u_s - u_i)^{\mathrm{T}} x < 0, \quad \forall s \in I \backslash \{i\}, \\ & (v_t - v_j)^{\mathrm{T}} x < 0, \quad \forall t \in J \backslash \{j\}, \end{aligned}$$

其中 $x \in \mathbf{R}^n$, 则有

$$U \dot{-} V = \mathrm{co}\{u_i - v_j | (L_{ij}) \text{ 有解}, i \in I, j \in J\}. \tag{5.3.9}$$

证明 不失一般性, 假设

$$u_s \neq u_t, \quad \forall s,t \in I, \quad s \neq t \text{ 和 } v_s \neq v_t, \quad \forall s,t \in J, \quad s \neq t.$$

记

$$\begin{aligned}
\Gamma_{ij} &= \{x \in \mathbf{R}^n | x \text{ 为 } (L_{ij}) \text{ 的解 }\}, \quad i \in I, j \in J, \\
U_{st} &= \{x \in \mathbf{R}^n | (u_s - u_t)^{\mathrm{T}} x = 0\}, \quad s,t \in I, \\
V_{st} &= \{x \in \mathbf{R}^n | (v_s - v_t)^{\mathrm{T}} x = 0\}, \quad s,t \in J, \\
I(x) &= \left\{ i \in I \,\middle|\, u_i^{\mathrm{T}} x = \max_{k \in I} u_k^{\mathrm{T}} x \right\}, \quad x \in \mathbf{R}^n, \\
J(x) &= \left\{ j \in J \,\middle|\, v_j^{\mathrm{T}} x = \max_{k \in J} v_k^{\mathrm{T}} x \right\}, \quad x \in \mathbf{R}^n.
\end{aligned}$$

下面证明

$$\operatorname{meas}\Big(\mathbf{R}^n \backslash \bigcup_{i \in I, j \in J} \Gamma_{ij}\Big) = 0. \tag{5.3.10}$$

设 $x \in \mathbf{R}^n \backslash \bigcup_{i \in I, j \in J} \Gamma_{ij}$, 则指标集 $I(x)$ 和 $J(x)$ 中至少有一个不是单点集, 否则存在 $i_0 \in I$ 和 $j_0 \in J$ 满足 $I(x) = \{i_0\}$ 和 $J(x) = \{j_0\}$, 于是

$$\begin{aligned}
(u_s - u_{i_0})^{\mathrm{T}} x < 0, \quad \forall s \in I \backslash \{i_0\}, \\
(v_t - v_{j_0})^{\mathrm{T}} x < 0, \quad \forall t \in J \backslash \{j_0\},
\end{aligned}$$

即 $x \in \Gamma_{i_0 j_0}$, 故 $x \in \bigcup_{i \in I, j \in J} \Gamma_{ij}$, 这与 $x \in \mathbf{R}^n \backslash \bigcup_{i \in I, j \in J} \Gamma_{ij}$ 矛盾. 不妨假设 $I(x)$ 不是单点集, 即存在 $s_0, t_0 \in I$, 使得

$$u_{s_0}^{\mathrm{T}} x = u_{t_0}^{\mathrm{T}} x = \max_{k \in I} u_k^{\mathrm{T}} x,$$

这意味着 $x \in U_{s_0 t_0}$, 于是有 $x \in \bigcup_{s,t \in I} U_{st}$, 从而可得

$$\mathbf{R}^n \backslash \bigcup_{i \in I, j \in J} \Gamma_{ij} \subset \Big(\bigcup_{s,t \in I} U_{st}\Big) \bigcup \Big(\bigcup_{s,t \in J} V_{st}\Big). \tag{5.3.11}$$

由假设 $u_s \neq u_t, \forall s,t \in I, s \neq t$ 和 $v_s \neq v_t, \forall s,t \in J, s \neq t$ 知, U_{st} 和 V_{st} 都是 $\mathbf{R}^n$ 中的 $n-1$ 维子空间, 测度均为零, 又因为 I 和 J 为有限指标集, 故有

$$\operatorname{meas}\Big(\Big(\bigcup_{s,t \in I} U_{st}\Big) \bigcup \Big(\bigcup_{s,t \in J} V_{st}\Big)\Big) = 0. \tag{5.3.12}$$

由式 (5.3.11) 和式 (5.3.12) 可得式 (5.3.10). 从 $x \in \Gamma_{ij}$ 不难看出

$$\left\{ u \in U \,\middle|\, u^{\mathrm{T}} x = \delta_U^*(x) \right\} = \{u_i\},$$

$$\left\{v \in V \left| v^{\mathrm{T}} x = \delta_V^* (x) \right.\right\} = \{v_j\},$$

且有

$$\nabla \delta^* (x | U) = u_i, \quad \nabla \delta^* (x | V) = v_j.$$

因此, $\delta^* (x | U)$ 和 $\delta^* (x | V)$ 在每个 Γ_{ij} 上可微, 从而在 $\bigcup\limits_{i \in I, j \in J} \Gamma_{ij}$ 上可微. 我们在 $U \dot{-} V$ 的定义中取 $T = \bigcup\limits_{i \in I, j \in J} \Gamma_{ij}$, 即可得式 (5.3.9). 定理得证.

注记 5.3.1　每个 (L_{ij}) 都是一个具有 $\mathrm{card} I + \mathrm{card} J - 2$ 个线性不等式的不等式组.

基于定理 5.3.1, 我们可通过确定系统对于每对指标 $i \in I$, $j \in J$ 的相容性来构造集合 $U \dot{-} V$.

定理 5.3.5　设 U 和 V 为定理 5.3.1 中给出的两集合, 给定一对指标 $i \in I$, $j \in J$, 定义线性不等式系统:

$$\begin{aligned} (\bar{L}_{ij}) \quad & (u_s - u_i)^{\mathrm{T}} x \leqslant 0, \quad \forall s \in I \backslash \{i\}, \\ & (v_t - v_j)^{\mathrm{T}} x < 0, \quad \forall t \in J \backslash \{j\}, \end{aligned}$$

其中 $x \in \mathbf{R}^n$, 则 $U \dot{-} V$ 具有如下形式:

$$U \dot{-} V = \mathrm{co} \left\{u_i - v_j \left| (\bar{L}_{ij}) \text{ 有非零解}, \ i \in I, j \in J \right.\right\}.$$

证明　给定一对指标 $i \in I$, $j \in J$, $(\bar{L}_i (U))$ 和 $(\bar{L}_i (V))$ 分别表示如下系统:

$$\begin{aligned} (\bar{L}_i (U)) \quad & (u_s - u_i)^{\mathrm{T}} x \leqslant 0, \quad \forall i \in I \backslash \{i\}, \\ (\bar{L}_j (V)) \quad & (v_s - v_j)^{\mathrm{T}} x \leqslant 0, \quad \forall j \in J \backslash \{j\}, \end{aligned}$$

其中 $x \in \mathbf{R}^n$. 注意到 U 是一个有限个点的凸包, 根据极大面的定义, 有

$$\begin{aligned} G_x(U) &= \mathrm{co} \left\{u_i \left| u_i^{\mathrm{T}} x = \max_{s \in I} u_s^{\mathrm{T}} x, i \in I \right.\right\} \\ &= \mathrm{co} \left\{u_i \left| (u_s - u_i)^{\mathrm{T}} x \leqslant 0, \forall s \in I \backslash \{i\}, i \in I \right.\right\} \\ &= \mathrm{co} \{u_i | x \text{ 为 } (\bar{L}_i (U)) \text{ 的解}, i \in I\}. \end{aligned}$$

类似可得

$$G_x (V) = \mathrm{co} \{v_j | x \text{ 为 } (\bar{L}_j (V)) \text{ 的解}, j \in J\}.$$

因为 $(\bar{L}_{ij})$ 是 $(\bar{L}_i (U))$ 和 $(\bar{L}_j (V))$ 的组合, 所以有

$$\begin{aligned} G_x (U) - G_x (V) &= \mathrm{co} \{u_i - u_j | x \text{ 是 } (\bar{L}_i (U)) \text{ 和 } (\bar{L}_j (V)) \text{ 的解}, i \in I, j \in J\} \\ &= \mathrm{co} \{u_i - u_j | x \text{ 是 } (\bar{L}_{ij}) \text{ 的解}, i \in I, j \in J\}, \end{aligned}$$

而且

$$\begin{aligned}U \dot{-} V &= \operatorname{clco} \bigcup_{x \neq 0} (G_x(U) - G_x(V)) \\ &= \operatorname{clco}\{u_i - u_j \mid \exists x \neq 0 \text{ 是 } (\bar{L}_{ij}) \text{ 的解}, i \in I, j \in J\} \\ &= \operatorname{co}\left\{u_i - u_j \,\middle|\, (\bar{L}_{ij}) \text{ 有非零解}, i \in I, j \in J\right\}.\end{aligned}$$

定理得证.

5.4 拟微分的代表元

从定义可以看出, 如果 $[\underline{\partial} f(x), \overline{\partial} f(x)]$ 是 $f(x)$ 的一个拟微分, 则对任意凸紧集 S, 集合 $[\underline{\partial} f(x) + S, \overline{\partial} f(x) - S]$ 也是 $f(x)$ 的一个拟微分, 因此拟微分不是唯一的. 本节讨论拟微分代表元的选取方法.

考虑下述集合对

$$\left[\bigcap_{[\underline{\partial} f(x), \overline{\partial} f(x)] \in Df(x)} (\underline{\partial} f(x) + \overline{\partial} f(x)), \bigcap_{[\underline{\partial} f(x), \overline{\partial} f(x)] \in Df(x)} (\overline{\partial} f(x) - \overline{\partial} f(x))\right], \quad (5.4.1)$$

集合对 (5.4.1) 被期望是 $f(x)$ 的一个拟微分, 如果是则可以作为拟微分代表元, 也称为拟微分核.

如果对任意满足关系

$$\underline{\partial} f(x) \subset \underline{\partial}^m f(x) \text{ 和 } \overline{\partial} f(x) \subset \overline{\partial}^m f(x)$$

的 $[\underline{\partial} f(x), \overline{\partial} f(x)] \in Df(x)$, 都有

$$[\underline{\partial} f(x), \overline{\partial} f(x)] = [\underline{\partial}^m f(x), \overline{\partial}^m f(x)],$$

则称 $[\underline{\partial}^m f(x), \overline{\partial}^m f(x)] \in Df(x)$ 为最小拟微分. 现已证明最小拟微分是存在的, 如果

$$[\underline{\partial} f(x), \overline{\partial} f(x)] \in Df(x),$$

那么存在一个最小拟微分

$$[\underline{\partial}^m f(x), \overline{\partial}^m f(x)] \in Df(x)$$

满足

$$\underline{\partial}^m f(x) \subset \underline{\partial} f(x), \quad \overline{\partial}^m f(x) \subset \overline{\partial} f(x).$$

可是最小拟微分并不唯一, 最小拟微分的任意平移仍是最小拟微分, 如果 $[A, B]$ 是一个最小拟微分, 那么对于任意单点集 $\{c\}$, $[A, B]$ 的平移 $[A + \{c\}, B - \{c\}]$ 仍是一个最小拟微分.

5.4.1 二维形式

对于二维形式, 在一定意义下, 最小拟微分是唯一的. 设 $f(x)$ 在 $\mathbf{R}^2$ 上存在拟微分, 任给两个最小拟微分

$$[\underline{\partial}_1^m f(x), \overline{\partial}_1^m f(x)] \text{ 和 } [\underline{\partial}_2^m f(x), \overline{\partial}_2^m f(x)],$$

那么存在 $c \in \mathbf{R}^2$(其中 c 依赖于 $[\underline{\partial}_1^m f(x), \overline{\partial}_1^m f(x)]$ 和 $[\underline{\partial}_2^m f(x), \overline{\partial}_2^m f(x)]$) 满足

$$[\underline{\partial}_2^m f(x), \overline{\partial}_2^m f(x)] = [\underline{\partial}_1^m f(x) + \{c\}, \overline{\partial}_1^m f(x) - \{c\}]. \tag{5.4.2}$$

下述定理说明在二维空间中拟微分核是存在的.

定理 5.4.1 设 $f(x)$ 是在 $\mathbf{R}^2$ 上的拟可微函数, $[\underline{\partial}_0^m f(x), \overline{\partial}_0^m f(x)]$ 是 $f(x)$ 在 x 的一个最小拟微分, 那么下述关系成立:

$$\bigcap_{[\underline{\partial} f(x), \overline{\partial} f(x)] \in Df(x)} (\underline{\partial} f(x) + \overline{\partial} f(x)) = \underline{\partial}_0^m f(x) + \overline{\partial}_0^m f(x), \tag{5.4.3a}$$

$$\bigcap_{[\underline{\partial} f(x), \overline{\partial} f(x)] \in Df(x)} (\overline{\partial} f(x) - \overline{\partial} f(x) = \overline{\partial}_0^m f(x) - \overline{\partial}_0^m f(x), \tag{5.4.3b}$$

进而有

$$\left[\bigcap_{[\underline{\partial} f(x), \overline{\partial} f(x)] \in Df(x)} (\underline{\partial} f(x) + \overline{\partial} f(x)), \bigcap_{[\underline{\partial} f(x), \overline{\partial} f(x)] \in Df(x)} (\overline{\partial} f(x) - \overline{\partial} f(x)\right] \in Df(x). \tag{5.4.4}$$

证明 设

$$[\underline{\partial} f(x), \overline{\partial} f(x)] \in Df(x),$$

根据最小拟微分的存在性可以得到, 存在一个 $f(x)$ 在 x 的最小拟微分

$$[\underline{\partial}^m f(x), \overline{\partial}^m f(x)],$$

满足

$$\underline{\partial}^m f(x) \subset \underline{\partial} f(x), \quad \overline{\partial}^m f(x) \subset \overline{\partial} f(x),$$

从而

$$\underline{\partial}^m f(x) + \overline{\partial}^m f(x) \subset \underline{\partial} f(x) + \overline{\partial} f(x), \tag{5.4.5a}$$

$$\overline{\partial}^m f(x) - \overline{\partial}^m f(x) \subset \overline{\partial} f(x) - \overline{\partial} f(x). \tag{5.4.5b}$$

注意到

$$[\underline{\partial}^m f(x), \overline{\partial}^m f(x)] \text{ 和 } [\underline{\partial}_0^m f(x), \overline{\partial}_0^m f(x)]$$

都是 $f(x)$ 在 x 的最小拟微分, 根据二维情形下最小拟微分的平移不变性, 存在 $c \in \mathbf{R}^2$, 使得最小拟微分 $[\underline{\partial}^m f(x), \overline{\partial}^m f(x)]$ 可表示成

$$[\underline{\partial}^m f(x), \overline{\partial}^m f(x)] = [\underline{\partial}_0^m f(x) + \{c\}, \overline{\partial}_0^m f(x) - \{c\}]. \tag{5.4.6}$$

由此推出

$$\underline{\partial}^m f(x) + \overline{\partial}^m f(x) = \underline{\partial}_0^m f(x) + \overline{\partial}_0^m f(x), \tag{5.4.7a}$$

$$\overline{\partial}^m f(x) - \overline{\partial}^m f(x) = \overline{\partial}_0^m f(x) - \overline{\partial}_0^m f(x). \tag{5.4.7b}$$

由式 (5.4.5a), 式 (5.4.5b), 式 (5.4.7a) 和式 (5.4.7b) 可以得到

$$\underline{\partial}_0^m f(x) + \overline{\partial}_0^m f(x) \subset \underline{\partial} f(x) + \overline{\partial} f(x), \tag{5.4.8a}$$

$$\overline{\partial}_0^m f(x) - \overline{\partial}_0^m f(x) \subset \overline{\partial} f(x) - \overline{\partial} f(x). \tag{5.4.8b}$$

对式 (5.4.8a) 和式 (5.4.8b) 右端关于所有拟微分取交, 有

$$\underline{\partial}_0^m f(x) + \overline{\partial}_0^m f(x) \subset \bigcap_{[\underline{\partial} f(x), \overline{\partial} f(x)] \in Df(x)} (\underline{\partial} f(x) + \overline{\partial} f(x)), \tag{5.4.9a}$$

$$\overline{\partial}_0^m f(x) - \overline{\partial}_0^m f(x) \subset \bigcap_{[\underline{\partial} f(x), \overline{\partial} f(x)] \in Df(x)} (\overline{\partial} f(x) - \overline{\partial} f(x)). \tag{5.4.9b}$$

另一方面, 由 $[\underline{\partial}_0^m f(x), \overline{\partial}_0^m f(x)] \in Df(x)$ 得

$$\bigcap_{[\underline{\partial} f(x), \overline{\partial} f(x)] \in Df(x)} (\underline{\partial} f(x) + \overline{\partial} f(x)) \subset \underline{\partial}_0^m f(x) + \overline{\partial}_0^m f(x), \tag{5.4.10a}$$

$$\bigcap_{[\underline{\partial} f(x), \overline{\partial} f(x)] \in Df(x)} (\overline{\partial} f(x) - \overline{\partial} f(x)) \subset \overline{\partial}_0^m f(x) - \overline{\partial}_0^m f(x). \tag{5.4.10b}$$

由式 (5.4.9a), 式 (5.4.9b), 式 (5.4.10a) 和式 (5.4.10b) 得

$$\bigcap_{[\underline{\partial} f(x), \overline{\partial} f(x)] \in Df(x)} (\underline{\partial} f(x) + \overline{\partial} f(x)) = \underline{\partial}_0^m f(x) + \overline{\partial}_0^m f(x), \tag{5.4.11a}$$

$$\bigcap_{[\underline{\partial} f(x), \overline{\partial} f(x)] \in Df(x)} (\overline{\partial} f(x) - \overline{\partial} f(x) = \overline{\partial}_0^m f(x) - \overline{\partial}_0^m f(x). \tag{5.4.11b}$$

注意到

$$[\underline{\partial}_0^m f(x), \overline{\partial}_0^m f(x)] \in Df(x) \text{ 和 } \overline{\partial}_0^m f(x)$$

为 $\mathbf{R}^2$ 中的凸紧集, 则有

$$[\underline{\partial}_0^m f(x) + \overline{\partial}_0^m f(x), \overline{\partial}_0^m f(x) - \overline{\partial}_0^m f(x)] \in Df(x). \tag{5.4.12}$$

式 (5.4.11a), 式 (5.4.11b) 和式 (5.4.12) 说明式 (5.4.4) 成立. 定理得证.

5.4.2 多维形式

定理 5.4.1 中, 式 (5.4.4) 之所以对二维拟可微函数成立, 主要是由于最小拟微分在二维情形下是相互平移的, 在 n 维 ($n \geqslant 3$) 的情况下最小拟微分由于不具有平移性质, 定理 5.4.1 的结论一般是不成立的. 但是有下面的结果:

$$\bigcap_{[\underline{\partial} f(x),\overline{\partial} f(x)]\in Df(x)} (\underline{\partial} f(x) + \overline{\partial} f(x)) = \bigcap_{[\underline{\partial}^m f(x),\overline{\partial}^m f(x)]\in D^m f(x)} (\underline{\partial}^m f(x) + \overline{\partial}^m f(x)), \tag{5.4.13a}$$

$$\bigcap_{[\underline{\partial} f(x),\overline{\partial} f(x)]\in Df(x)} (\overline{\partial} f(x) - \overline{\partial} f(x)) = \bigcap_{[\underline{\partial}^m f(x),\overline{\partial}^m f(x)]\in D^m f(x)} (\overline{\partial}^m f(x) - \overline{\partial}^m f(x)), \tag{5.4.13b}$$

其中 $D^m f(x)$ 表示 $f(x)$ 在 x 的所有最小拟微分所组成的集合.

定理 5.4.2　假设 $f(x)$ 是 $\mathbf{R}^n$ 上的拟可微函数, 且在 x 点存在一个拟微分 $[\underline{\partial}_0 f(x),\overline{\partial}_0 f(x)]$ 满足

$$\underline{\partial}_0 f(x) \dot{-} (-\overline{\partial}_0 f(x)) = \underline{\partial}_0 f(x) - (-\overline{\partial}_0 f(x)), \tag{5.4.14}$$

则有

$$\bigcap_{[\underline{\partial} f(x),\overline{\partial} f(x)]\in Df(x)} (\underline{\partial} f(x) + \overline{\partial} f(x)) = \underline{\partial}_0 f(x) + \overline{\partial}_0 f(x), \tag{5.4.15a}$$

$$\bigcap_{[\underline{\partial} f(x),\overline{\partial} f(x)]\in Df(x)} (\overline{\partial} f(x) - \overline{\partial} f(x) = \overline{\partial}_0 f(x) - \overline{\partial}_0 f(x), \tag{5.4.15b}$$

进而有

$$\left[\bigcap_{[\underline{\partial} f(x),\overline{\partial} f(x)]\in Df(x)} (\underline{\partial} f(x) + \overline{\partial} f(x)), \bigcap_{[\underline{\partial} f(x),\overline{\partial} f(x)]\in Df(x)} (\overline{\partial} f(x) - \overline{\partial} f(x)\right] \in Df(x). \tag{5.4.16}$$

证明　设

$$[\underline{\partial} f(x),\overline{\partial} f(x)] \in Df(x),$$

由式 (5.4.14) 得到

$$\begin{aligned}\underline{\partial}_0 f(x) + \overline{\partial}_0 f(x) &= \underline{\partial}_0 f(x) \dot{-} (-\overline{\partial}_0 f(x)) \\ &= \underline{\partial} f(x) \dot{-} (-\overline{\partial} f(x)) \\ &\subset \underline{\partial} f(x) + \overline{\partial} f(x).\end{aligned} \tag{5.4.17}$$

根据拟微分的定义, 由

$$[U,V] \in Df(x),$$

则有

$$[U+V, U-V] \in Df(x),$$

因此

$$[\underline{\partial} f(x)+\overline{\partial} f(x), \overline{\partial} f(x)-\overline{\partial} f(x)]$$

和

$$[\underline{\partial}_0 f(x)+\overline{\partial}_0 f(x), \overline{\partial}_0 f(x)-\overline{\partial}_0 f(x)]$$

都是 $f(x)$ 在 x 点的拟微分. 根据拟微分定义,

$$\begin{aligned} &\max_{u\in\underline{\partial} f(x)+\overline{\partial} f(x)} u^{\mathrm{T}}y + \min_{v\in\overline{\partial} f(x)-\overline{\partial} f(x)} v^{\mathrm{T}}y \\ =& \max_{u\in\underline{\partial}_0 f(x)+\overline{\partial}_0 f(x)} u^{\mathrm{T}}y + \min_{v\in\overline{\partial}_0 f(x)-\overline{\partial}_0 f(x)} v^{\mathrm{T}}y, \quad \forall y\in\mathbf{R}^n. \end{aligned} \tag{5.4.18}$$

根据式 (5.4.17), 有

$$\max_{u\in\underline{\partial} f(x)+\overline{\partial} f(x)} u^{\mathrm{T}}y \geqslant \max_{u\in\underline{\partial}_0 f(x)+\overline{\partial}_0 f(x)} u^{\mathrm{T}}y, \quad \forall y\in\mathbf{R}^n. \tag{5.4.19}$$

结合式 (5.4.18) 和式 (5.4.19) 得

$$\min_{v\in\overline{\partial} f(x)-\overline{\partial} f(x)} v^{\mathrm{T}}y \leqslant \min_{v\in\overline{\partial}_0 f(x)-\overline{\partial}_0 f(x)} v^{\mathrm{T}}y, \quad \forall y\in\mathbf{R}^n. \tag{5.4.20}$$

同理, 式 (5.4.20) 等价于

$$\max_{v\in\overline{\partial} f(x)-\overline{\partial} f(x)} v^{\mathrm{T}}y \geqslant \min_{v\in\overline{\partial}_0 f(x)-\overline{\partial}_0 f(x)} v^{\mathrm{T}}y, \quad \forall y\in\mathbf{R}^n. \tag{5.4.21}$$

由式 (5.4.21) 及凸紧集和其支撑函数的关系得

$$\overline{\partial}_0 f(x)-\overline{\partial}_0 f(x)\subset\overline{\partial} f(x)-\overline{\partial} f(x). \tag{5.4.22}$$

注意到式 (5.4.17) 和式 (5.4.22) 对任意的 $[\underline{\partial} f(x), \overline{\partial} f(x)]\in Df(x)$ 都是成立的, 分别在式 (5.4.17) 和式 (5.4.22) 的右边对所有拟微分取交得

$$\underline{\partial}_0 f(x)+\overline{\partial}_0 f(x)\subset\bigcap_{[\underline{\partial} f(x),\overline{\partial} f(x)]\in Df(x)}(\underline{\partial} f(x)+\overline{\partial} f(x)), \tag{5.4.23a}$$

$$\overline{\partial}_0 f(x)-\overline{\partial}_0 f(x)\subset\bigcap_{[\underline{\partial} f(x),\overline{\partial} f(x)]\in Df(x)}(\overline{\partial} f(x)-\overline{\partial} f(x)). \tag{5.4.23b}$$

另一方面, 由

$$[\underline{\partial}_0 f(x), \overline{\partial}_0 f(x)]\in Df(x)$$

可以推出

$$\bigcap_{[\underline{\partial} f(x),\overline{\partial} f(x)]\in Df(x)}(\underline{\partial} f(x)+\overline{\partial} f(x))\subset\underline{\partial}_0 f(x)+\overline{\partial}_0 f(x), \tag{5.4.24a}$$

$$\bigcap_{[\underline{\partial} f(x),\overline{\partial} f(x)]\in Df(x)} (\overline{\partial} f(x) - \overline{\partial} f(x) \subset \overline{\partial}_0 f(x) - \overline{\partial}_0 f(x). \tag{5.4.24b}$$

结合式 (5.4.23a) 和式 (5.4.24b) 得到式 (5.4.15a). 同理, 结合式 (5.4.23b) 和式 (5.4.24b), 并注意到关系

$$[\underline{\partial}_0 f(x) + \overline{\partial}_0 f(x), \overline{\partial}_0 f(x) - \overline{\partial}_0 f(x)] \in Df(x).$$

得式 (5.4.15b). 命题得证.

定义 5.4.1　设 X_1, X_2 为 $\mathbf{R}^n$ 中子空间, 如果 $X_1 + X_2 = \mathbf{R}^n$, 且对于任意 $x_1 \in X_1$ 和 $x_2 \in X_2$ 都有 $x_1^{\mathrm{T}} x_2 = 0$, 则称 X_1 和 X_2 是正交补子空间.

定理 5.4.3　设 X_1 和 X_2 是 $\mathbf{R}^n$ 的正交补子空间, 如果 $A \subset X_1$, $B \subset X_2$, 那么 $A \dot{-} B = A - B$.

证明　根据式 (5.3.7), 有

$$A \dot{-} B = \partial(\delta^*(x\,|A) - \delta^*(x\,|B))|_{x=0}. \tag{5.4.25}$$

由于 X_1 和 X_2 为正交补子空间, 则任意 $x \in \mathbf{R}^n$ 都可以表示为 $x = x_1 + x_2$, 其中 $x_1 \in X_1$, $x_2 \in X_2$. 显然

$$\delta^*(x\,|A) - \delta^*(x\,|B) = \delta^*(x_1\,|A) - \delta^*(x_2\,|B), \tag{5.4.26}$$

注意到 $x_1 \in X_1$, $x_2 \in X_2$, X_1 和 X_2 是正交补子空间. 在这种情况下, Clarke 广义梯度形式下的链式法则正好是等式关系. 由 (5.4.26) 可得

$$\begin{aligned} \partial(\delta^*(x\,|A) - \delta^*(x\,|B))|_{x=0} &= \partial(\delta^*(x_1\,|A) - \delta^*(x_2\,|B))|_{x_1=x_2=0} \\ &= \partial\delta^*(x_1\,|A)|_{x_1=0} - \partial\delta^*(x_2\,|B)|_{x_2=0}. \end{aligned} \tag{5.4.27}$$

对于支撑函数, Clarke 广义梯度为凸函数的次微分, 在式 (5.4.27) 中, 令

$$A = \partial\delta^*(x_1\,|A)\,|_{x_1=0} \;\text{ 和 }\; B = \partial\delta^*(x_2\,|B)\,|_{x_2=0}\,,$$

利用支撑函数在 0 点次微分公式, 得

$$\partial\delta^*(x_1\,|A)|_{x_1=0} - \partial\delta^*(x_2\,|B)| = A - B. \tag{5.4.28}$$

由式 (5.4.28) 和式 (5.4.25) 即得

$$A \dot{-} B = A - B.$$

定理得证.

引理 5.4.1　设 $f(x)$ 为 $\mathbf{R}^n$ 上的拟可微函数, $[A, B]$, $[U, V] \in Df(x)$, 那么 V 有如下的形式:

$$V = A \dot{-} (U - B). \tag{5.4.29}$$

证明 显然

$$\max_{u\in A} u^{\mathrm{T}}y - \max_{v\in -B} v^{\mathrm{T}}y = \max_{u\in U} u^{\mathrm{T}}y - \max_{v\in -V} v^{\mathrm{T}}y, \quad \forall y \in \mathbf{R}^n,$$

故

$$\max_{u\in A} u^{\mathrm{T}}y - \max_{v\in U-B} v^{\mathrm{T}}y = -\max_{v\in -V} v^{\mathrm{T}}y, \quad \forall y \in \mathbf{R}^n. \tag{5.4.30}$$

在 $y=0$ 处取 Clarke 广义梯度, 式 (5.4.30) 变成

$$\partial\left(\max_{u\in A} u^{\mathrm{T}}y - \max_{v\in U-B} v^{\mathrm{T}}y\right)\Big|_{y=0} = \partial\left(-\max_{v\in -V} v^{\mathrm{T}}y\right)\Big|_{y=0}. \tag{5.4.31}$$

根据 Demyanov 差的定义和式 (5.4.31), 得

$$A\dot{-}(U-B) = -(-V),$$

即式 (5.4.29) 成立. 引理得证.

定理 5.4.4 设 $f(x)$ 是 $\mathbf{R}^n$ 上的拟可微函数, 且 $[\underline{\partial}f(x),\overline{\partial}f(x)]\in Df(x)$, 如果存在 $W\subset\mathbf{R}^n$, 使得

$$[\underline{\partial}f(x)\dot{-}(-\overline{\partial}f(x)), W]\in Df(x),$$

那么

$$W = \underline{\partial}f(x)\dot{-}\{(\underline{\partial}f(x)\dot{-}(-\overline{\partial}f(x))) - \overline{\partial}f(x)\}.$$

证明 在引理 5.4.1 中, 取

$$[A,B] = [\underline{\partial}f(x),\overline{\partial}f(x)],$$

$$[U,V] = [\underline{\partial}f(x)\dot{-}(-\overline{\partial}f(x)), W],$$

则有

$$W = A\dot{-}(U-B) = \underline{\partial}f(x)\dot{-}\{(\underline{\partial}f(x)\dot{-}(-\overline{\partial}f(x))) - \overline{\partial}f(x)\}.$$

命题得证.

定理 5.4.5 设 $f(x)$ 是 $\mathbf{R}^n$ 上的拟可微函数, 如果存在 $[A,B]\in Df(x)$ 满足 $A\dot{-}(-B) = A-(-B)$, 那么对任意的拟微分 $[\underline{\partial}f(x),\overline{\partial}f(x)]\in Df(x)$, 集合对

$$[\underline{\partial}f(x)\dot{-}(-\overline{\partial}f(x)), \underline{\partial}f(x)\dot{-}\{(\underline{\partial}f(x)\dot{-}(-\overline{\partial}f(x))) - \overline{\partial}f(x)\}] \tag{5.4.32}$$

是 $f(x)$ 在 x 点的拟微分.

证明 根据定理 5.4.2,

$$[A\dot{-}(-B), B-B] = [\underline{\partial}f(x)\dot{-}(-\overline{\partial}f(x)), B-B]$$

是拟微分. 根据定理 5.4.4, 由

$$[\underline{\partial} f(x) \dot{-} (-\overline{\partial} f(x)), B - B] \in Df(x)$$

可得

$$B - B = \underline{\partial} f(x) \dot{-} \{(\underline{\partial} f(x) \dot{-} (-\overline{\partial} f(x))) - \overline{\partial} f(x)\}.$$

这说明式 (5.4.32) 是拟微分. 命题得证.

注意到多面体的 Demyanov 差分和 Minkowski 差分是多面体, 我们有如下的推论.

推论 5.4.1　假设存在 $[A, B] \in Df(x)$ 满足 $A \dot{-} (-B) = A - (-B)$, 另有多面体 $[U, V] \in Df(x)$, 那么拟微分 (5.4.13) 是一对多面体.

设 $[\underline{\partial} f(x), \overline{\partial} f(x)] \in Df(x)$, $[U, V]$ 一对凸紧集, 那么 $[U, V] \in Df(x)$ 当且仅当下面等式成立:

$$\underline{\partial} f(x) - V = U - \overline{\partial} f(x). \tag{5.5.33}$$

设 $[U, V]$ 为式 (5.4.32) 和式 (5.4.33) 中的集合, 则有

$$\begin{aligned} &\underline{\partial} f(x) - \{\underline{\partial} f(x) \dot{-} [(\underline{\partial} f(x) \dot{-} (-\overline{\partial} f(x))) - \overline{\partial} f(x)]\} \\ =& \{\underline{\partial} f(x) \dot{-} (-\overline{\partial} f(x))\} - \overline{\partial} f(x). \end{aligned} \tag{5.4.34}$$

由此, 对于任意的 $[\underline{\partial} f(x), \overline{\partial} f(x)] \in Df(x)$, 式 (5.4.32) 是拟微分当且仅当式 (5.4.34) 成立. 这样, 给定一个拟微分 $[\underline{\partial} f(x), \overline{\partial} f(x)]$, 我们首先检验式 (5.4.32) 是否是一个拟微分, 或等价地, 检验式 (5.4.34) 是否成立, 如果成立, 取式 (5.4.32) 中不依赖 $[\underline{\partial} f(x), \overline{\partial} f(x)]$ 的集合作为拟微分等价类的代表.

5.5　矩阵空间上凸紧集的差

本节介绍矩阵空间上凸紧集的差, 它可用于通过拟微分计算 Clarke 广义 Jacobi 和 B 微分.

5.5.1　定义及性质

设 $W = W_1 \times \cdots \times W_m$, 其中 $W_1, \cdots, W_m \subset \mathbf{R}^n$. 矩阵集合 W 的支持函数 $\delta^*(x|W)$ 定义为

$$\delta^*(x\,|W) = (\delta^*(x\,|W_1), \cdots, \delta^*(x\,|W_m))^{\mathrm{T}}.$$

W 的最大面和最小面分别用 $G_x(W)$ 和 $\tilde{G}_x(W)$ 表示, 定义为

$$G_x(W) = \{\,(w_1, \cdots, w_m)^{\mathrm{T}} \,|\, w_i^{\mathrm{T}} x = \max_{w \in W_i} w^{\mathrm{T}} x, w_i \in W_i, i = 1, \cdots, m\},$$

$$\tilde{G}_x(W) = \{(w_1,\cdots,w_m)^{\mathrm{T}} \big| w_i^{\mathrm{T}} x = \min_{w \in W_i} w^{\mathrm{T}} x, w_i \in W_i, i = 1,\cdots,m\},$$

$G_x(W)$ 和 $\tilde{G}_x(W)$ 中的元素分别用 $g_x(W)$ 和 $\tilde{g}_x(W)$ 表示.

设 U_i 和 V_i, $i = 1,\cdots,m$ 是 $\mathbf{R}^n$ 上的凸紧集, 记

$$[U, V] = [U_1, V_1] \times \cdots \times [U_m, V_m]. \tag{5.5.1}$$

给定满测度集 T(定义见 5.3 节), 使得在每一点 $x \in T$ 处, $\delta^*(x\,|U)$ 和 $\delta^*(x\,|V)$ 的广义 Jacobi 存在, 定义矩阵集合 U 和 V 的差 $U \dot{-} V$ 定义如下:

$$U \dot{-} V = \mathrm{cl}\{J\delta^*(x\,|U) - J\delta^*(x\,|V)|\, x \in T\}, \tag{5.5.2}$$

其中 J 代表 Jacobi. 注意到 $x \in T$, 可以推出 $G_x(U)$ 和 $\tilde{G}_x(V)$ 都是单点集, 因此, $U \dot{-} V$ 可以等价地表示为

$$U \dot{-} V = \mathrm{cl}\{g_x(U) - \tilde{g}_x(V)|\, x \in T\}.$$

首先, 我们要说明式 (5.5.2) 定义的合理性, 即 $U \dot{-} V$ 不依赖于集合 T 的选择. 假设 T_{UV} 是 $J\delta^*(x|U)$ 和 $J\delta^*(x|V)$ 存在的点的集合, $T \subset T_{UV}$ 是一个满测度集, 于是

$$T_{UV} \subset \mathbf{R}^n = \mathrm{cl}T.$$

记

$$W_1 = \mathrm{cl}\,\{J\delta^*(x\,|U) - J\delta^*(x\,|V)|\, x \in T_{UV}\},$$

$$W_2 = \mathrm{cl}\,\{J\delta^*(x\,|U) - J\delta^*(x\,|V)|\, x \in T\},$$

显然 $W_2 \subset W_1$. 另一方面, 设 $x \in T_{UV}$, 则有 $T_{UV} \subset \mathrm{cl}T$, 于是存在序列 $\{x_k\}$ 满足 $x_k \in T$, $x_k \to x$. 既然 Jacobi $J\delta^*(x\,|U)$ 存在, 那么 $J\delta^*(x_k\,|U) \to J\delta^*(x\,|U)$. 同理, $J\delta^*(x_k\,|V) \to J\delta^*(x\,|V)$, 因此

$$J\delta^*(x\,|U) - J\delta^*(x\,|V) \in W_2,$$

即 $W_1 \subset W_2$, 故 $W_1 = W_2$. 这样如果 $T_1, T_2 \subset T_{UV}$, T_1 和 T_2 是满测度的, 有

$$\mathrm{cl}\,\{J\delta^*(x\,|U) - J\delta^*(x\,|V)|\, x \in T_1\} = \mathrm{cl}\,\{J\delta^*(x\,|U) - J\delta^*(x\,|V)|\, x \in T_2\}.$$

这说明, 式 (5.5.2) 不依赖于 T 的选择, 即 $U \dot{-} V$ 是适定的.

当 $m = 1$(即 $U, V \subset \mathbf{R}^n$), 集合 $\mathrm{co}(U \dot{-} V)$ 是 Demyanov 差, 不难得出

$$U \dot{-} V \subset (U_1 \dot{-} V_1) \times \cdots \times (U_m \dot{-} V_m).$$

但上式只是包含关系, 一般情况下等式并不成立.

下面介绍 $\mathbf{R}^{m\times n}$ 中凸紧集的另一种差. 设 U 和 V 由式 (5.5.1) 中给出, 集合的差 $U\dot{-}V$ 定义为

$$U\dot{-}V=\bigcup_{x\neq 0}[G_x(U)-G_x(V)].$$

容易验证集合 $U\dot{-}V$ 是闭的.

显然, 当 $m=1$ 时, $\mathrm{co}(U\dot{-}V)$ 是 U 和 V 的 Rubinov 差, 事实上

$$U\dot{-}V\subset(U_1\dot{-}V_1)\times\cdots\times(U_m\dot{-}V_m).$$

5.5.2 拟微分表示 B 微分和广义 Jacobi

下面给出基于集合的差, 利用拟微分表示 B 微分和 Clarke 广义 Jacobi.

设 $F(x)$ 是 $\mathbf{R}^n$ 到 $\mathbf{R}^m$ 上的向量拟可微函数, 集合 $T\subset\mathbf{R}^n$ 称为关于拟微分 $[\underline{\partial}F(x),\overline{\partial}F(x)]$ 满足性质 (ε), 如果下述条件成立:

(i) 集合 $\mathbf{R}^n\backslash T$ 的测度为零;

(ii) 对于任意的 $g\in T$, 集合 $G_g(\underline{\partial}F(x))$ 和 $\tilde{G}_g(\overline{\partial}F(x))$ 是单点集.

设 $X\subset\mathbf{R}^n$ 为开集, 点 $x\in X$, 定义 X 到 $\mathbf{R}^m$ 的一族函数类 $M(x)$ 如下: 如果 $F(x)\in M$, 则满足以下条件:

(a) F 在 x 的邻域 $B(x,\delta)$ 是局部 Lipschitz 的, 这说明 D_F 关于 $B(x,\delta)$ 是满测度的, 即 $B(x,\delta)\backslash D_F$ 的测度为零;

(b) F 在 x 是拟可微的;

(c) 存在关于 $B(x,\delta)$ 是满测度的 $Q\subset D_f$ 和关于拟微分 $[\underline{\partial}F(x),\overline{\partial}F(x)]$ 满足性质 (ε) 的 T, 且由条件 $g_k\to g$, $t_k\to 0^+$, $x_k=x+t_kg_k\in Q$, $g\in T$ 可以得到

$$JF(x_k)\to g_g(\underline{\partial}F(x))+\tilde{g}_g(\overline{\partial}F(x)).$$

下面的定理给出了 B 微分和拟微分之间的关系.

定理 5.5.1 给定一点 $x\in\mathbf{R}^n$, 如果 $F(x)$ 属于 $\mathbf{R}^n$ 到 $\mathbf{R}^m$ 上的函数 $M(x)$, 那么下述关系成立:

$$\underline{\partial}F(x)\dot{-}(-\overline{\partial}F(x))\subset\partial_BF(x). \tag{5.5.3}$$

证明 取集合 $Q\subset B(x,\delta)\bigcap D_F$ 满足

$$\mathrm{meas}(B(x,\delta)\backslash Q)=0, \tag{5.5.4}$$

另取 $T\subset\mathbf{R}^n$ 满足 $\mathrm{meas}(\mathbf{R}^n\backslash T)=0$ 及性质 (ε). 记

$$\partial_TF(x)=\mathrm{cl}\{\xi|\,x_k=x+t_kg_k\in Q,g_k\to g\in T,t_k\to 0,JF(x_k)\to\xi\}.$$

设 $v \in \partial_T F(x)$, 那么存在序列 $\{g_k\}$ 和 $\{t_k\}$ 满足

$$g_k \to g, \quad t_k \to 0^+, \quad x_k = x + t_k g_k \in Q, \quad g \in T, \quad JF(x_k) \to v.$$

另一方面, 由 $F(x) \in M(x)$ 可得

$$JF(x_k) \to g_g(\underline{\partial} F(x)) + \tilde{g}_g(\overline{\partial} F(x)),$$

即

$$v = g_g(\underline{\partial} F(x)) + \tilde{g}_g(\overline{\partial} F(x)).$$

这样有

$$\begin{aligned} \partial_T F(x) &\subset \mathrm{cl}\{\, g_g(\underline{\partial} F(x)) + \tilde{g}_g(\overline{\partial} F(x)) \,|\, g \in T\} \\ &= \underline{\partial} F(x) \dot{-} (-\overline{\partial} F(X)). \end{aligned} \tag{5.5.5}$$

假设 $g \in \mathbf{R}^n$, $t_k \to 0^+$, $\varepsilon_k \to 0^+$, 由式 (5.5.4) 可得

$$Q \bigcap (x + t_k B(g, \varepsilon_k)) \neq \varnothing,$$

那么存在 $g_k \in B(g, \varepsilon_k)$ 满足

$$x + t_k g_k \in Q. \tag{5.5.6}$$

因此, 能找到一个序列 $\{g_k\}$ 满足 $g_k \to g$ 和式 (5.5.6) 成立. 取 $g \in T$, 考虑元素

$$v = g_g(\underline{\partial} F(x)) + \tilde{g}_g(\overline{\partial} F(x)),$$

序列 $\{g_k\}$ 和 $\{t_k\}$ 满足 $g_k \to g$, $t_k \to 0^+$, 式 (5.5.6) 成立. 根据 $M(x)$ 的定义, $JF(x) \to v$ 是成立的, 即

$$v \in \mathrm{cl}\{\, \xi |\, x_k = x + t_k g_k \in Q, g_k \to g \in T, t_k \to 0^+, JF(x_k) \to \xi\} = \partial_T F(x).$$

由此得

$$\underline{\partial} F(x) \dot{-} (-\overline{\partial} F(x)) = \partial_T F(x). \tag{5.5.7}$$

根据式 (5.5.5) 和式 (5.5.7) 得

$$\underline{\partial} F(x) \dot{-} (-\overline{\partial} F(x)) = \partial_T F(x).$$

注意到 $\partial_B F(x)$ 是闭集, 有

$$\partial_T F(x) \subset \partial_B F(x),$$

式 (5.5.3) 成立. 命题得证.

推论 5.5.1　设 $F(x)$ 为 $\mathbf{R}^n$ 到 $\mathbf{R}^m$ 上的函数, 给定一点 $x \in \mathbf{R}^n$, 如果 $F(x) \in M(x)$, 则有

$$\mathrm{co}(\underline{\partial}F(x) \dot{-} (-\overline{\partial}F(x))) \subset \partial F(x).$$

下面讨论函数族 $M(x)$.

命题 5.5.1　设 $F(x)$ 为 $\mathbf{R}^n$ 到 $\mathbf{R}^m$ 上的函数, 则 $F(x)$ 属于 $M(x)$ 当且仅当 $F(x)$ 的每个分量属于 $M(x)$.

证明　设 $F(x) = (f_1(x), \cdots, f_m(x))^{\mathrm{T}}$, 其中 $f_i(x)$, $i = 1, \cdots, m$ 为 $\mathbf{R}^n$ 上的实值函数, 显然 $f_i(x) \in M(x)$. 为了证明 $F \in M(x)$, 我们只需考虑条件 (c). 假设条件 (c) 对每个 $f_i(x)$ 都成立, 即存在一个满测度的集合 $Q_i \subset D_{f_i}$ 且满足关系:

$$g_k \to g, \quad t_k \to 0^+, \quad x_k = x + t_k g_k \in Q_i, \quad g \in T_i,$$

使得

$$\nabla f_i(x_k) \to g_g(\underline{\partial} f_i(x)) + \tilde{g}_g(\overline{\partial} f_i(x)).$$

取

$$Q = \bigcap_{i=1}^{m} Q_i, \quad T = \bigcap_{i=1}^{m} T_i,$$

$$g_k \to g, \quad t_k \to 0^+, \quad x_k = x + t_k g_k \in Q_i, \quad g \in T,$$

由此得

$$JF(x_k) \to g_g(\underline{\partial} F(x)) + \tilde{g}_g(\overline{\partial} F(x)).$$

这说明条件 (c) 对 $F(x)$ 成立. 定理得证.

根据命题 5.5.1, 我们得到如下的结果.

命题 5.5.2　设 $F(x)$ 为 $\mathbf{R}^n$ 到 $\mathbf{R}^m$ 上的函数, 如果 $F(x)$ 在点 x 的某邻域内是凸的, 那么 $F(x) \in M(x)$.

命题 5.5.3　设 $F_i(x)$ 为 $\mathbf{R}^n$ 到 $\mathbf{R}^m$ 上的函数, $F_i \in M(x)$, $i = 1, \cdots, l$, $G(x)$ 是连续可微函数. 那么复合函数

$$G(F_1(x), \cdots, F_l(x)) \in M(x).$$

命题 5.5.4　假设 X 是 $\mathbf{R}^n$ 中的一个开集, $F_i(x)$, $i = 1, \cdots, l$ 为 X 到 $\mathbf{R}^m$ 上的向量函数, 且属于 $M(x)$, 那么函数 $\max\limits_{1 \leqslant i \leqslant l} F_i(x)$ 和 $\min\limits_{1 \leqslant i \leqslant l} F_i(x)$ 属于 $M(x)$.

定理 5.5.2　假设 U 和 V 如式 (5.5.1) 所给, 那么以下公式成立:

$$U \dot{-} V = \partial_B(\delta^*(\cdot|U) - \delta^*(\cdot|V))(0). \tag{5.5.8}$$

证明 由于 $\delta^*(\cdot\,|U)$ 和 $\delta^*(\cdot\,|V)$ 都是凸函数, 由性质 5.5.2 和 5.5.3, 对任意的 $x\in\mathbf{R}^n$, $\delta^*(\cdot\,|U)-\delta^*(\cdot\,|V)$ 包含在函数族 $M(x)$ 中. 记

$$G(\cdot)=\delta^*(\cdot\,|U)-\delta^*(\cdot\,|V),$$

那么

$$\underline{\partial}G(0)=U,\quad \overline{\partial}G(0)=-V$$

是 $G(x)$ 在 $x=0$ 的一个拟微分, 由定理 5.5.1 得

$$U\dot{-}V\subset\partial_B(\delta^*(\cdot\,|U)-\delta^*(\cdot\,|V))(0). \tag{5.5.9}$$

下面证明

$$U\dot{-}V\supset\partial_B(\delta^*(\cdot\,|U)-\delta^*(\cdot\,|V))(0).$$

记

$$D=\mathrm{cl}\{\,w|\,x_k\to 0,J(\delta^*(\cdot\,|U)-\delta^*(\cdot\,|V))(x_k)\to w,x_k\in T_{UV}\},$$

其中

$$\partial_B(\delta^*(\cdot\,|U)-\delta^*(\cdot\,|V))(0)=D.$$

设 $w\in D$, 则序列 $\{x_k\}$ 满足

$$x_k\to 0,\quad J(\delta^*(\cdot\,|U)-\delta^*(\cdot\,|V))(x_k)\to w.$$

由

$$J(\delta^*(\cdot\,|U)-\delta^*(\cdot\,|U))(x_k)=J\delta^*(x_k\,|U)-J\delta^*(x_k\,|V),$$

可得

$$w\in\mathrm{cl}\{\,J\delta^*(x\,|U)-J\delta^*(x\,|V)\,|\,x\in T_{UV}\}=U\dot{-}V,$$

由此得

$$D\subset U\dot{-}V,$$

进而有

$$\partial_B(\delta^*(\cdot\,|U)-\delta^*(\cdot\,|V))(0)\subset U\dot{-}V. \tag{5.5.10}$$

结合式 (5.5.9) 和式 (5.5.10) 可以推出式 (5.5.8). 命题得证.

定理 5.5.3 设 $F(x)$ 为 $\mathbf{R}^n$ 到 $\mathbf{R}^m$ 上函数, 在 x 处是拟可微的, $[\underline{\partial}F(x),\overline{\partial}F(x)]$ 为其一个拟微分, 则有

$$\partial_B(F'(x;y)\,)|_{y=0}=\underline{\partial}F(x)\dot{-}(-\overline{\partial}F(x)). \tag{5.5.11}$$

证明　注意到

$$F'(x;y) = \delta^*(y\,|\underline{\partial}F(x)) - \delta^*(y\,|\overline{\partial}F(x)),$$

由式 (5.5.8) 得式 (5.5.11) 成立. 命题得证.

尽管拟微分不是唯一的, 但式 (5.5.11) 说明, 次微分与负超微分的差是唯一确定的.

定理 5.5.4　假设 $\mathbf{R}^n$ 到 $\mathbf{R}^m$ 上的函数 $F(x)$ 在 x 的一个邻域内分片连续可微, 那么式 (5.5.3) 成立.

除具有 $M(x)$ 的条件 (a), (b) 外, 函数族 $\tilde{M}(x)$ 还要求满足下述条件:

(d) 存在 $B(x,\delta)$ 上的满测度的集合 $Q \subset D_F$ 和拟微分 $[\underline{\partial}F(x), \overline{\partial}F(x)]$, 使得由关系式

$$g_k \to g, \quad t_k \downarrow 0^+, \quad x_k = x + t_k g_k \in D_F, \quad JF(x_k) \to v$$

可以得到

$$v \in G_g(\underline{\partial}F(x)) + \tilde{G}_g(\overline{\partial}F(x)).$$

定理 5.5.5　设 $\mathbf{R}^n$ 到 $\mathbf{R}^m$ 上的函数 $F(x)$ 包含在 $\tilde{M}(x)$ 中, $[\underline{\partial}F(x), \overline{\partial}F(x)]$ 是条件 (d) 中的一个拟微分, 那么下述关系成立:

$$\partial_B F(x) \subset \underline{\partial}F(x) \dot{-} (-\overline{\partial}F(x)). \tag{5.5.12}$$

证明　设

$$\xi \in \partial_B F(x) = \{\, v |\, x_k \in D_F, x_k \to x, JF(x_k) \to v\},$$

序列 $\{v_k\}$ 满足

$$x_k \in D_F, \quad x_k \to x, \quad JF(x_k) \to \xi.$$

记

$$\alpha_k = \|x_k - x\|, \quad g_k = \frac{x_k - x}{\alpha_k}.$$

不失一般性, 假设 $\alpha_k \downarrow 0$, $\lim_{k\to\infty} g_k = g$, 根据性质 (d), 有

$$\xi \in G_g(\underline{\partial}F(x) + \tilde{G}_g(\overline{\partial}F(x)) = G_g(\underline{\partial}F(x)) - G_g(-\overline{\partial}F(x)),$$

由此得

$$\partial_B F(x) \subset \underline{\partial}F(x) \dot{-} (-\overline{\partial}F(x)).$$

定理得证.

命题 5.5.5　设 $F(x)$ 为 $\mathbf{R}^n$ 到 $\mathbf{R}^m$ 上的向量值函数, 则 $F(x)$ 属于函数类 $\tilde{M}(x)$ 当且仅当它的每个分量都包含在 $\tilde{M}(x)$ 中.

命题 5.5.6 设 $F(x)$ 为 $\mathbf{R}^n$ 到 $\mathbf{R}^m$ 上的向量值函数, 如果 $F(x)$ 在 x 的一个邻域内是凸的, 那么

$$F(x) \in \tilde{M}(x).$$

命题 5.5.7 假设 $F_i(x)$ 为 $\mathbf{R}^n$ 到 $\mathbf{R}^m$ 上的向量值函数, $F_i(x) \in M(x)$, $i = 1, \cdots, l$, $G(y)$ 为 $\mathbf{R}^l$ 上的连续可微函数, 那么 $G(F_1(x), \cdots, F_l(x)) \in \tilde{M}(x)$.

命题 5.5.8 假设 X 是 $\mathbf{R}^n$ 中的一个开集, $F_i(x)$, $i = 1, \cdots, l$ 为 X 到 $\mathbf{R}^m$ 上的函数, 且属于 $\tilde{M}(x)$, 那么 $\max\limits_{1 \leqslant i \leqslant l} F_i(x)$ 和 $\min\limits_{1 \leqslant i \leqslant l} F_i(x)$ 属于函数类 $\tilde{M}(x)$ 中.

既然 $\tilde{M}(x) \subset M(x)$, 则有

$$\begin{aligned} \underline{\partial} F(x) \dot{-} (-\overline{\partial} F(x)) &= \partial_B F(x) \\ &\subset \underline{\partial} F(x) \ddot{-} (-\overline{\partial} F(x)), \quad \forall F \in \tilde{M}(x), \end{aligned} \tag{5.5.13}$$

由式 (5.5.13), 我们有如下性质:

命题 5.5.9 假设 $F(x)$ 为 $\mathbf{R}^n$ 到 $\mathbf{R}^m$ 上的函数,

$$F(x) \in \tilde{M}(x),$$

$$\underline{\partial} F(x) \dot{-} (-\overline{\partial} F(x)) = \underline{\partial} F(x) \ddot{-} (-\overline{\partial} F(x)),$$

则有

$$\begin{aligned} \partial_B F(x) &= \underline{\partial} F(x) \dot{-} (-\overline{\partial} F(x)) \\ &= \partial_B F(x) \subset \underline{\partial} F(x) \ddot{-} (-\overline{\partial} F(x)). \end{aligned} \tag{5.5.14}$$

5.5.3 多面体差的公式

下面我们针对多面体 U 和 V 给出 $U \dot{-} V$ 和 $U \ddot{-} V$ 的表达式.

定理 5.5.6 设 $U, V \in \mathbf{R}^{m \times n}$, 如式 (5.5.1) 所给出,

$$U_i = \mathrm{co}\{u_{ij} | j \in J_i\}, \quad V_i = \mathrm{co}\{v_{ik} | k \in K_i\}, \quad i = 1, \cdots, m,$$

其中 $u_{ij}, v_{ik} \in \mathbf{R}^n$, J_i, K_i 是有限指标集, 不失一般性, 假设

$$u_{is} \neq v_{it}, \quad \forall s, t \in J_i, \quad s \neq t, \quad v_{is} \neq v_{it}, \quad \forall s, t \in K_i, \quad s \neq t, \quad i = 1, \cdots, m.$$

给定一对指标集

$$j_1 \in J_1, \cdots, j_m \in J_m \text{ 和 } k_1 \in K_1, \cdots, k_m \in K_m,$$

令线性不等式系统如下:

$$
\begin{aligned}
(L(j_1,\cdots,j_m;k_1,\cdots,k_m))\quad & (u_{1_s}-u_{1_{j_1}})^{\mathrm{T}}y<0,\quad \forall s\in J_1\backslash\{j_1\},\\
& \cdots\cdots\cdots\cdots\\
& (u_{ms}-u_{mj_m})^{\mathrm{T}}y<0,\quad \forall s\in J_m\backslash\{j_m\},\\
& (v_{1t}-v_{1k_1})^{\mathrm{T}}y<0,\quad \forall t\in K_1\backslash\{k_1\},\\
& \cdots\cdots\cdots\cdots\\
& (v_{mt}-v_{mk_m})^{\mathrm{T}}y<0,\quad \forall t\in K_m\backslash\{k_m\},
\end{aligned}
$$

其中 $y\in\mathbf{R}^n$, 那么

$$
U\dot{-}V=\left\{(u_{1j_1}-v_{1k_1},\cdots,u_{mj_m}-v_{mk_m})^{\mathrm{T}}\,\middle|\,\begin{array}{l}(L(j_1,\cdots,j_m;k_1,\cdots,k_m))\text{ 有解}\\ j_p\in J_p,k_p\in K_p,1\leqslant p\leqslant m\end{array}\right\}. \tag{5.5.15}
$$

证明　记

$$
\begin{aligned}
&\Gamma(j_1,\cdots,j_m;k_1,\cdots,k_m)=\left\{y\in\mathbf{R}^n\,\middle|\,\begin{array}{l}y\text{ 是 }(L(j_1,\cdots,j_m;k_1,\cdots,k_m))\text{的解}\\ j_p\in J_p,k_p\in K_p,p=1,\cdots,m\end{array}\right\},\\
&U_{ijs}=\{y\in\mathbf{R}^n\,|(u_{ij}-u_{is})^{\mathrm{T}}y=0\},\quad j,s\in J_i,i=1,\cdots,m,\\
&V_{ikt}=\{y\in\mathbf{R}^n\,|(v_{ik}-v_{it})^{\mathrm{T}}y=0\},\quad k,t\in K_i,i=1,\cdots,m,\\
&J_i(y)=\{j\in J_i|\,u_{ij}^{\mathrm{T}}y=\max_{s\in J_i}u_{is}^{\mathrm{T}}y\},\quad y\in\mathbf{R}^n,i=1,\cdots,m,\\
&K_i(y)=\{k\in K_i|\,v_{ik}^{\mathrm{T}}y=\max_{t\in K_i}v_{it}^{\mathrm{T}}y\},\quad y\in\mathbf{R}^n,i=1,\cdots,m,
\end{aligned}
$$

以下面证明

$$
\operatorname{meas}\Big(\mathbf{R}^n\backslash\bigcup_{j_p\in J_p,k_p\in K_p}\Gamma(j_1,\cdots,j_m;k_1,\cdots,k_m)\Big)=0. \tag{5.5.16}
$$

设

$$
\overline{y}\notin\mathbf{R}^n\backslash\bigcup_{j_p\in J_p,k_p\in K_p}\Gamma(j_1,\cdots,j_m;k_1,\cdots,k_m)),
$$

即

$$
\overline{y}\notin\mathbf{R}^n\backslash\bigcup_{j_p\in J_p,k_p\in K_p}\Gamma(j_1,\cdots,j_m;k_1,\cdots,k_m)),
$$

则有

$$
\overline{y}\notin\Gamma(j_1,\cdots,j_m;k_1,\cdots,k_m),\quad \forall j_p\in J_p,\quad k_p\in K_p,\quad p=1,\cdots,m.
$$

因此, 存在一个指标 $1\leqslant i\leqslant m$, 使得指标集 $J_i(\overline{y})$ 和 $K_i(\overline{y})$ 中至少有一个不是单点集, 否则 $J_i(\overline{y})$ 和 $K_i(\overline{y})$, $i=1,\cdots,m$ 都是单点集, 记

$$
J_1(\overline{y})=\{j_1\},\cdots,J_m(\overline{y})=\{j_m\},
$$

$$K_1(\overline{y}) = \{k_1\}, \cdots, K_m(\overline{y}) = \{k_m\}.$$

这样, $y = \overline{y}$ 是系统 $(L(j_1, \cdots, j_m; k_1, \cdots, k_m))$ 的解, 这与 $\overline{y} \notin \Gamma(j_1, \cdots, j_m; k_1, \cdots, k_m))$ 矛盾. 不妨设存在一个指标 $1 \leqslant i_0 \leqslant m$, 使得 $J_{i_0}(\overline{y})$ 不是单点集, 于是存在 $j_0, s_0 \in J_{i_0}(\overline{y})$, $j_0 \neq s_0$, 使得 $\overline{y} \in U_{i_0 j_0 s_0}$, 由此得

$$\overline{y} \in \bigcup_{1 \leqslant i \leqslant m} \Big(\bigcup_{j,s \in j_i} U_{ijs} \Big),$$

进而推出

$$\mathbf{R}^n \backslash \bigcup_{j_p \in J_p, k_p \in K_p} \Gamma(j_1, \cdots, j_m; k_1, \cdots, k_m) \subset \bigcup_{1 \leqslant i \leqslant m} \Big(\Big(\bigcup_{j,s \in J_i} U_{ijs} \Big) \bigcup \Big(\bigcup_{k,t \in K_i} V_{ikt} \Big) \Big). \tag{5.5.17}$$

根据假设条件 $u_{is} \neq u_{it}$, $\forall s, t \in J_i$, $s \neq t$ 和 $v_{is} \neq v_{it}$, $\forall s, t \in K_i$, $s \neq t$, $i = 1, \cdots, m$, 得 U_{ijs} 和 V_{ikt} 是 $\mathbf{R}^n$ 中的 $n-1$ 维子空间, 所以它们的测度为零, 又因为 J_i, K_i 是有限指标集, 所以有

$$\text{meas}\Big\{ \bigcup_{1 \leqslant i \leqslant m} \Big[\Big(\bigcup_{j,s \in J_i} U_{ijs} \Big) \cup \Big(\bigcup_{k,t \in K_i} V_{ikt} \Big) \Big] \Big\} = 0, \tag{5.5.18}$$

由式 (5.5.18) 和式 (5.5.17) 可以得式 (5.5.16). 不难看出, 由 $\overline{y} \in \Gamma(j_1, \cdots, j_m; k_1, \cdots, k_m)$ 可得

$$G_{\overline{y}}(U_i) = \{u_{ij_i}\}, \quad G_{\overline{y}}(V_i) = \{v_{ij_i}\}, \quad i = 1, \cdots, m,$$

因此, $\delta^*(y\,|U_i)$ 和 $\delta^*(y\,|V_i)$ 在 $\overline{y}$ 处可微, 且有 $\nabla\delta^*(\overline{y}\,|U_i) = u_{ij_1}$ 和 $\nabla\delta^*(\overline{y}\,|V_i) = v_{ij_1}, i = 1, \cdots, m$. 进而有 $\delta^*(y\,|U_i)$ 和 $\delta^*(y\,|V_i)$ 在 $\overline{y}$ 处可微, 具有

$$J\delta^*(\overline{y}\,|U) = (u_{1j_1}, \cdots, u_{mj_m})^{\mathrm{T}}$$

和

$$J\delta^*(\overline{y}\,|V) = (v_{1k_1}, \cdots, v_{mk_m})^{\mathrm{T}}.$$

因此, $\delta^*(y\,|U)$ 和 $\delta^*(y\,|V)$ 在 $\Gamma(j_1, \cdots, j_m; k_1, \cdots, k_m)$ 内可微, 对于 $j_1 \in J_1, \cdots, j_m \in J_m$ 和 $k_1 \in K_1, \cdots, k_m \in K_m$, 在 $U \dot{-} V$ 定义中取

$$T = \bigcup_{j_p \in J_p, k_p \in K_p} \Gamma(j_1, \cdots, j_m; k_1, \cdots, k_m),$$

即得式 (5.5.15). 命题得证.

注 5.5.1 系统 $(L(j_1, \cdots, j_m; k_1, \cdots, k_m))$ 是有 n 个变量的严格线性不等式组.

定理 5.5.6 说明, 可通过确定系统 $(L(j_1,\cdots,j_m;k_1,\cdots,k_m))$ 的相容性得到 $U\dot{-}V$, 事实上, 给定 $j_1\in J_1,\cdots,j_m\in J_m$ 和 $k_1\in K_1,\cdots,k_m\in K_m$, 如果系统 $(L(j_1,\cdots,j_m;k_1,\cdots,k_m))$ 是相容的, 那么

$$(u_{1j_1}-v_{1k_1},\cdots,u_{mj_m}-v_{mk_m})^{\mathrm{T}}\in U\dot{-}V,$$

否则

$$(u_{1j_1}-v_{1k_1},\cdots,u_{mj_m}-v_{mk_m})^{\mathrm{T}}\notin U\dot{-}V.$$

定理 5.5.7　假设 $U,V\in\mathbf{R}^{m\times n}$ 如定理 5.5.6 所给出, 对于两组指标集 $j_1\in J_1,\cdots,j_m\in J_m$ 和 $k_1\in K_1,\cdots,k_m\in K_m$, 构造如下的线性不等式系统:

$$\begin{aligned}
(M(j_1,\cdots,j_m;k_1,\cdots,k_m))\quad &(u_{1s}-u_{1j_1})^{\mathrm{T}}y\leqslant 0,\qquad \forall s\in J_1\backslash\{j_1\},\\
&\cdots\cdots\cdots\cdots\\
&(u_{ms}-u_{mj_m})^{\mathrm{T}}y\leqslant 0,\quad \forall s\in J_m\backslash\{j_m\},\\
&(v_{1t}-v_{1k_1})^{\mathrm{T}}y\leqslant 0,\qquad \forall t\in K_1\backslash\{k_1\},\\
&\cdots\cdots\cdots\cdots\\
&(v_{mt}-v_{mk_m})^{\mathrm{T}}y\leqslant 0,\quad \forall t\in K_m\backslash\{k_m\},
\end{aligned}$$

其中 $y\in\mathbf{R}^n$, 那么

$$U\dot{-}V=\left\{(u_{1j_1}-v_{1k_1},\cdots,u_{mj_m}-v_{mk_m})^{\mathrm{T}}\;\middle|\;\begin{array}{l}(M(j_1,\cdots,j_m;k_1,\cdots,k_m))\text{有非零解},\\ j_p\in J_p,k_p\in K_p,1\leqslant p\leqslant m\end{array}\right\}.\tag{5.5.19}$$

证明　给定一对指标集 $j_1\in J_1,\cdots,j_m\in J_m$ 和 $k_1\in K_1,\cdots,k_m\in K_m$, 构造两个线性不等式系统:

$$\begin{aligned}
(U_{j_1,\cdots,j_m})\quad &(u_{1s}-u_{1j_1})^{\mathrm{T}}y\leqslant 0,\qquad \forall s\in J_1\backslash\{j_1\},\\
&\cdots\cdots\cdots\cdots\\
&(u_{ms}-u_{mj_m})^{\mathrm{T}}y\leqslant 0,\quad \forall s\in J_m\backslash\{j_m\};\\
(V_{k_1,\cdots,k_m})\quad &(v_{1t}-v_{1k_1})^{\mathrm{T}}y\leqslant 0,\qquad \forall t\in K_1\backslash\{k_1\},\\
&\cdots\cdots\cdots\cdots\\
&(v_{mt}-v_{mk_m})^{\mathrm{T}}y\leqslant 0,\quad \forall t\in K_m\backslash\{k_m\},
\end{aligned}$$

其中 $y\in R^n$. 根据最大面的定义, 有

$$G_x(U)=\{(u_{1j_1},\cdots,u_{mj_m})^{\mathrm{T}}|\,u_{ij_i}^{\mathrm{T}}x=\max_{s\in J_i}u_{is}^{\mathrm{T}}x,\quad j_i\in J_i,\quad i=1,\cdots,m\}$$

$$=\left\{(u_{1j_1},\cdots,u_{mj_m})^{\mathrm{T}}\ \middle|\ \begin{array}{l}(u_{1s}-u_{1j_1})^{\mathrm{T}}x\leqslant 0,\quad \forall s\in J_1\backslash\{j_1\}\\ \cdots\cdots\cdots\cdots\\ (u_{ms}-u_{mj_m})^{\mathrm{T}}x\leqslant 0,\quad \forall s\in J_m\backslash\{j_m\}\\ j_i\in J_i, i=1,\cdots,m\end{array}\right\}$$

$$=\{(u_{1j_1},\cdots,u_{mj_m})^{\mathrm{T}}|\, x \text{ 为 } (U_{j_1,\cdots,j_m})\text{的解},\ j_p\in J_p, p=1,\cdots,m\},$$

同理可得

$$G_x(V)=\{(v_{1k_1},\cdots,v_{mk_m})^{\mathrm{T}}|\, x \text{ 为 } (V_{k_1\cdots k_m}) \text{ 的解},\ k_p\in K_p, p=1,\cdots,m\}.$$

注意到, 系统 $(M(j_1,\cdots,j_m;k_1,\cdots,k_m))$ 是系统 $(U_{j_1,\cdots,j_m})$ 和 $(V_{k_1,\cdots,k_m})$ 的组合, 则有

$$\begin{aligned}&G_x(U)-G_x(V)\\ =&\left\{(u_{1j_1},\cdots,u_{mj_m})^{\mathrm{T}}-(v_{1k_1},\cdots,v_{mk_m})^{\mathrm{T}}\ \middle|\ \begin{array}{l}x \text{ 解 } (U_{j_1,\cdots,j_m}),(V_{k_1,\cdots,k_m}),\\ j_p\in J_p,k_p\in K_p,p=1,\cdots,m\end{array}\right\}\\ =&\left\{(u_{1j_1}-v_{1k_1},\cdots,u_{mj_m}-v_{mk_m})^{\mathrm{T}}\ \middle|\ \begin{array}{l}x \text{ 解 } (M(j_1,\cdots,j_m;k_1,\cdots,k_m)),\\ j_p\in J_p,k_p\in K_p,1\leqslant p\leqslant m\end{array}\right\},\end{aligned}$$

进而有

$$\begin{aligned}U\dot{-}V=&\bigcup_{x\neq 0}[G_x(U)-G_x(V)]\\ =&\left\{(u_{1j_1}-v_{1k_1},\cdots,u_{mj_m}-v_{mk_m})^{\mathrm{T}}\ \middle|\ \begin{array}{l}x\neq 0 \text{ 是 } (M(j_1,\cdots,j_m;k_1,\cdots,k_m)) \text{ 的解},\\ j_p\in J_p,k_p\in K_p,1\leqslant p\leqslant m\end{array}\right\}\\ =&\left\{(u_{1j_1}-v_{1k_1},\cdots,u_{mj_m}-v_{mk_m})^{\mathrm{T}}\ \middle|\ \begin{array}{l}(M(j_1,\cdots,j_m;k_1,\cdots,k_m))\\ \text{有非零解},\\ j_p\in J_p,k_p\in K_p,1\leqslant p\leqslant m\end{array}\right\},\end{aligned}$$

即式 (5.5.19) 成立. 命题得证.

显然, 每个系统 $(M(j_1,\cdots,j_m;k_1,\cdots,k_m))$ 有 $\sum\limits_{i=1}^{m}(\mathrm{card}J_i+\mathrm{card}K_i-2)$ 个不等式.

第 6 章　最优性条件

本章讨论约束非光滑优化的最优性条件. 所谓最优性条件就是对一个优化问题最优点特征的刻画, 包括充分性条件和必要性条件两种, 但在最优化研究中人们往往对必要性条件关注多一些.

6.1　凸规划的最优性条件

所谓凸规划是指凸函数在凸集上的极小化问题. 由于凸规划具有较一般非凸规划更好的性质, 特别是在局部极小与整体极小的等价性以及必要性条件与充分性条件关系等方面有显著的特征, 因此备受人们关注.

6.1.1　一般约束情形

考虑下述优化问题:

$$\begin{aligned}&\min f(x),\\ &\text{s.t. } x \in S,\end{aligned} \tag{6.1.1}$$

其中 $f(x)$ 为 $\mathbf{R}^n$ 上的凸函数, $S \subset \mathbf{R}^n$ 为凸集.

定理 6.1.1　问题 (6.1.1) 的局部极小点一定是全局极小点, 进一步, 它的极小点集为 S 中的凸集.

证明　设 $x^* \in S$ 是问题 (6.1.1) 的局部极小点, 则存在 $r > 0$, 使得

$$f(x^*) \leqslant f(x), \quad \forall x \in B(x^*, r) \cap S.$$

对任意 $x \in S$ 且 $x \notin B(x^*, r)$, 令

$$t = \frac{r}{\|x - x^*\|}, \quad x_t = (1-t)x^* + tx,$$

显然

$$0 < t < 1, \quad x_t \in B(x^*, r) \cap S.$$

根据假设及 $f(x)$ 的凸性得

$$f(x^*) \leqslant f(x_t) \leqslant (1-t)f(x^*) + tf(x),$$

于是 $tf(x^*) \leqslant tf(x)$, 即 $f(x^*) \leqslant f(x)$, 故 x^* 为问题 (6.1.1) 的全局极小点.

另一方面, 问题 (6.1.1) 的最优解集可利用它的任意一个最优解 x^* 来表示:

$$S \cap \{x \in \mathbf{R}^n | f(x) \leqslant f(x^*)\}. \tag{6.1.2}$$

注意到, 凸函数的水平集是凸集, 集合 (6.1.2) 为两个凸集的交, 因此是凸集, 即问题 (6.1.1) 的解集为凸集. 定理得证.

定理 6.1.1 说明, 对于凸规划, 局部最优等同于全局最优, 因此, 对凸规划问题不再区分局部最优解与全局最优解.

定理 6.1.2 在问题 (6.1.1) 中, 只假设 $f(x)$ 为 $\mathbf{R}^n$ 上的局部 Lipschitz 函数, S 为 $\mathbf{R}^n$ 中任意子集, 如果 $x \in S$ 为问题 (6.1.1) 的最优解, 则存在常数 $\alpha > 0$ 和 $\delta > 0$, 使得当 $c \geqslant \alpha$ 时, 有

$$f(x) + cd_S(x) \leqslant f(y) + cd_S(y), \quad \forall y \in B(x, \delta), \tag{6.1.3}$$

其中 $d_S(x)$ 为点 x 到集合 S 的距离.

证明 根据 $f(x)$ 的局部 Lipschitz 性质, 存在 $\delta' > 0$, $\alpha > 0$, 使得当 $c \geqslant \alpha$ 时

$$|f(x_1) - f(x_2)| \leqslant c\|x_1 - x_2\|, \quad \forall x_1, x_2 \in B(x, \delta'). \tag{6.1.4}$$

取 $\varepsilon \in (0,1)$ 充分小及 $\delta \in (0, \delta'/3)$, 则对于 $y \in B(x, \delta)$, 存在 $y_\varepsilon \in S$, 使得

$$\|y - y_\varepsilon\| \leqslant (1+\varepsilon) d_S(y).$$

由于

$$\|x - y_\varepsilon\| \leqslant \delta < \delta'/3,$$

推导得

$$\begin{aligned}\|x - y_\varepsilon\| &\leqslant \|x - y\| + \|y - y_\varepsilon\| \\ &\leqslant \|x - y\| + (1+\varepsilon) d_S(y) \\ &\leqslant (2+\varepsilon)\|x - y\| < \delta'.\end{aligned}$$

因此, 由式 (6.1.4) 得

$$\begin{aligned}f(y_\varepsilon) &\leqslant f(y) + c(1+\varepsilon) d_S(y) \\ &\leqslant f(y) + c\,\|y - y_\varepsilon\| \\ &\leqslant f(y) + c(1+\varepsilon) d_S(y) \\ &\leqslant f(y) + c(1+\varepsilon) d_S(y).\end{aligned}$$

另一方面, 由 $y_\varepsilon \in S$ 和 $f(x) \leqslant f(y_\varepsilon)$ 以及 x 为极小点, 则有

$$\begin{aligned}f(x) &= f(x) + cd_S(x) \\ &\leqslant f(y) + c(1+\varepsilon) d_S(y),\end{aligned}$$

再由 ε 的任意性即得式 (6.1.3). 定理得证.

上述定理说明, 对于目标函数为局部 Lipschitz 的约束最优化问题 $\min\limits_{x\in S} f(x)$ 可等价的转化为无约束问题

$$\min f(x) + cd_S(x),$$

其中 c 为大于某一个门槛值 α 的常数即可.

定理 6.1.3 设 $x^* \in S$, 对于优化问题 (6.1.1), 下述结论等价:

(1) x^* 是问题 (6.1.1) 的最优解;

(2) $f'(x^*; y - x^*) \geqslant 0, \forall y \in S$;

(3) $f'(x^*; d) \geqslant 0, \forall d \in T_S(x^*)$;

(4) $0 \in \partial f(x^*) + N_S(x^*)$.

证明 (1)⇒ (2)⇒ (3) 假设结论 (1) 成立, 即 x^* 是问题 (6.1.1) 的最优解. 对任意 $y \in S$, 根据集合 S 的凸性有

$$x^* + t(y - x^*) \in S, \quad \forall t \in [0, 1].$$

由 x^* 是最优解, 则有

$$\frac{f(x^* + t(y - x^*)) - f(x^*)}{t} \geqslant 0, \quad \forall t \in [0, 1],$$

在上式中令 $t \to 0^+$, 得

$$f'(x^*; y - x^*) \geqslant 0,$$

故结论 (2) 成立.

假设 (2) 成立, 在 (2) 中令 $d = y - x^*$, 再考虑方向导数的正齐次性, 得

$$f'(x; d) \geqslant 0, \quad \forall d \in \operatorname{cone}(S - x^*),$$

而 $\operatorname{cone}(S - x^*)$ 的闭包恰好为 $T_S(x^*)$, 注意到方向导数 $f'(x; d)$ 关于 d 是连续的, 所以

$$f'(x; d) \geqslant 0, \quad \forall d \in T_S(x^*).$$

(3)⇒(1) 假设结论 (3) 成立, 对任意 $y \in S$, 显然

$$y - x^* \in T_S(x^*),$$

则有

$$0 \leqslant f'(x^*; y - x^*).$$

另一方面, 由 $f(x)$ 的凸性知

$$f'(x^*; y - x^*) \leqslant f(y) - f(x^*),$$

于是

$$0 \leqslant f(y) - f(x^*),$$

x^* 是 $f(x)$ 的最优解, 结论 (1) 成立.

(1) $\Leftrightarrow$ (4) 由于结论 (1) 成立, 根据定理 6.1.1, 存在常数 $c > 0$, 使得 x^* 是函数 $f(x) + cd_S(x)$ 的极小点. 注意到 $f(x) + cd_S(x)$ 是凸函数, 根据凸函数极值条件及 $\partial d_S(x^*) = N_S(x^*)$, 则有

$$\begin{aligned} 0 \in \partial(f(x^*) + d_S(x^*)) &= \partial f(x^*) + \partial d_S(x^*) \\ &= \partial f(x^*) + N_S(x^*), \end{aligned}$$

结论 (4) 成立. 如果结论 (4) 成立, 则

$$0 \in \partial(f(x^*) + d_S(x^*)),$$

x^* 是 $f(x) + d_S(x)$ 在 $\mathbf{R}^n$ 上的极小值, 即

$$\begin{aligned} f(x^*) &= f(x^*) + d_S(x^*) \\ &\leqslant f(x) + d_S(x), \quad \forall x \in \mathbf{R}^n. \end{aligned} \tag{6.1.5}$$

在式 (6.1.5) 中考虑 $x \in S$, 得

$$f(x^*) \leqslant f(x), \quad \forall x \in S,$$

结论 (1) 成立. 定理得证.

定理 6.1.3 给出了约束凸规划最优解的刻画, 结果具有一般性, 主要适用于约束没有明确解析表达式问题.

6.1.2 不等式约束的 Fritz John 最优性条件

下面考虑问题 (6.1.1) 约束集合 S 由不等式给出, 通过引入 Lagrange 乘子, 建立相应的最优性条件.

考虑下述不等式约束问题:

$$\begin{aligned} &\min f(x), \\ &\text{s.t. } g_i(x) \leqslant 0, \quad i = 1, \cdots, m, \end{aligned} \tag{6.1.6}$$

其中 $f(x)$, $g_i(x)$, $i = 1, \cdots, m$ 均为 $\mathbf{R}^n$ 上的凸函数.

定理 6.1.4 (Fritz John 条件) 设 $x^* \in \mathbf{R}^n$ 为问题 (6.1.6) 的最优解, 则存在一组不全为零的常数 $\lambda_0, \lambda_1, \cdots, \lambda_m \geqslant 0$ 满足 $\sum\limits_{i=0}^{m} \lambda_i = 1$, 使得

$$0 \in \lambda_0 \partial f(x^*) + \sum_{i=1}^{m} \lambda_i \partial g_i(x^*), \tag{6.1.7}$$

$$\lambda_i g_i(x^*) = 0. \tag{6.1.8}$$

证明　引入下述函数:

$$F(x) = \max\{f(x) - f(x^*), g_1(x), \cdots, g_m(x)\}. \tag{6.1.9}$$

不难验证

$$F(x) \geqslant 0, \quad \forall x \in \mathbf{R}^n,$$

而 $F(x^*) = 0$, 故 x^* 为 $F(x)$ 在 $\mathbf{R}^n$ 上的极小值点, 又 $F(x)$ 为凸函数, 根据无约束优化最优性条件, 有 $0 \in \partial F(x^*)$. 根据次微分运算法则有

$$\partial F(x^*) = \mathrm{co}\Big\{\partial f(x^*) \bigcup \big(\bigcup_{i \in I(x^*)} \partial g_i(x^*)\big)\Big\}, \tag{6.1.10}$$

其中

$$I(x^*) = \{i | g_i(x^*) = 0\}.$$

注意到, 式 (6.1.10) 右端项可表示为

$$\Big\{\lambda_0 \xi + \sum_{i \in I(x^*)} \lambda_i \xi_i | \xi \in \partial f(x^*), \xi_i \in \partial g_i(x^*), \lambda_i \geqslant 0, i \in I(x^*), \sum_{i \in I(x^*) \cup \{0\}} \lambda_i = 1\Big\},$$

于是

$$0 \in \lambda_0 \partial f(x^*) + \sum_{i \in I(x^*)} \lambda_i \partial g_i(x^*), \tag{6.1.11}$$

$$\lambda_i \geqslant 0, \quad i \in I(x^*) \bigcup \{0\}, \qquad \sum_{i \in I(x^*) \cup \{0\}} \lambda_i = 1. \tag{6.1.12}$$

令

$$\lambda_i = 0, \quad i \in \{1, \cdots, m\} \setminus I(x^*),$$

由式 (6.1.11) 和式 (6.1.12) 即得式 (6.1.7) 和式 (6.1.8). 定理得证.

式 (6.1.7) 和式 (6.1.8) 称为不等式约束凸规划的 Fritz John 最优性条件, $\lambda_i, i = 0, 1, \cdots, m$ 称为 Lagrange 乘子, $\lambda_i g_i(x^*) = 0, i = 0, 1, \cdots, m$ 称为互补松弛条件. 满足式 (6.1.7) 和式 (6.1.8) 的点称为优化问题 (6.1.6) 的 Fritz John 稳定点. 当 $f(x)$, $g_i(x)$, $i = 1, \cdots, m$ 为光滑凸函数时, 定理 6.1.4 即为通常光滑凸规划的 Fritz John 条件.

在式 (6.1.7) 中, 如果 $\lambda_0 \neq 0$, 此时可假设 $\lambda_0 = 1$, 为保证 $\lambda_0 \neq 0$, 我们引入下述约束品性.

约束品性 6.1.1(Slater 约束品性)　存在 $\hat{x} \in \mathbf{R}^n$, 使得 $g_i(\hat{x}) < 0, i = 1, \cdots, m$.

Slater 约束品性的含义为问题 (6.1.6) 可行域中有内点.

定理 6.1.5 (Kuhn-Tucker 条件) 设 $x^* \in \mathbf{R}^n$ 为问题 (6.1.6) 的最优解, 如果约束品性 6.1.1 成立, 则存在一组常数 $\lambda_1, \cdots, \lambda_m \geqslant 0$, 使得

$$0 \in \partial f(x^*) + \sum_{i=1}^{m} \lambda_i \partial g_i(x^*), \tag{6.1.13}$$

$$\lambda_i g_i(x^*) = 0, \quad i = 1, \cdots, m. \tag{6.1.14}$$

证明 根据定理 6.1.4, 式 (6.1.7) 和式 (6.1.8) 成立, 为证明式 (6.1.13) 和式 (6.1.14), 只需证明式 (6.1.7) 中 $\lambda_0 \neq 0$. 用反证法. 假设 $\lambda_0 = 0$, 定义 Lagrange 函数

$$L(x) = \sum_{i=1}^{m} \lambda_i g_i(x),$$

其中 $\lambda_i \geqslant 0, i = 1, \cdots, m$. 显然, $L(x)$ 是 $\mathbf{R}^n$ 上的凸函数, $L(x^*) = 0$, 由式 (6.1.7) 及 $\lambda_0 = 0$ 得 $0 \in \partial L(x^*)$. 根据无约束凸优化的最优性条件, x^* 是函数 $L(x)$ 的极小值点. 另一方面, $\lambda_i \geqslant 0, i = 1, \cdots, m$ 不全为零, 根据约束品性 6.1.1 有

$$L(\hat{x}) = \sum_{i=1}^{m} \lambda_i \partial g_i(\hat{x}) < 0,$$

与假设矛盾. 定理得证.

式 (6.1.13) 和式 (6.1.14) 称为凸规划 (6.1.6) 的 Kuhn-Tucker 条件, 满足式 (6.1.3) 和式 (6.1.4) 的点称为问题 (6.1.6) 的 Kuhn-Tucker 点. 当然, 也可以在其他约束品性下建立问题 (6.1.6) 的 Kuhn-Tucker 必要条件.

6.1.3 线性等式约束的情形

考虑下述含线性等式约束的凸规划问题:

$$\begin{aligned} &\min f(x) \\ &\text{s.t. } a_i^{\mathrm{T}} x = b_i, \quad i = 1, \cdots, p, \end{aligned} \tag{6.1.15}$$

其中 $f(x)$ 为 $\mathbf{R}^n$ 上的凸函数, $a_i, i = 1, \cdots, p$ 为 $\mathbf{R}^n$ 中的向量, $b_i, i = 1, \cdots, p$ 为常数.

这里需要指出的是在凸规划中等式约束只能考虑线性函数, 而非一般的凸函数, 这是因为当等式约束为非线性凸函数时, 约束集合已不再是凸集, 因而对应的优化问题也就不是凸规划了. 下述定理建立了问题 (6.1.15) 的 Kuhn-Tucker 必要条件.

定理 6.1.6 (Kuhn-Tucker 必要条件)　设 x^* 为问题 (6.1.15) 的最优解, 则存在常数 $\mu_1, \cdots, \mu_p$, 使得下式成立

$$0 \in \partial f(x^*) + \sum_{i=1}^{p} \mu_i a_i. \tag{6.1.16}$$

证明　记

$$D = \{x \in \mathbf{R}^n | a_i^{\mathrm{T}} x = b, i = 1, \cdots, p\}.$$

由于 x^* 为最优解, 根据定理 6.1.2, 则有

$$0 \in \partial f(x^*) + N_D(x^*).$$

再由法锥定义可得

$$N_D(x^*) = \left\{ \sum_{i=1}^{p} \mu_i a_i | \mu_i \in \mathbf{R}, i = 1, \cdots, p \right\},$$

故式 (6.1.16) 成立. 定理得证.

定理 6.1.6 说明, 对于线性等式约束的凸规划问题可以直接得到 Kuhn-Tucker 条件, 而不需要附加某种约束品性.

6.1.4　含等式和不等式约束的情形

考虑下述优化问题:

$$\begin{aligned} &\min f(x), \\ &\text{s.t. } g_i(x) \leqslant 0, \quad i = 1, \cdots, m, \\ &\qquad a_j^{\mathrm{T}} x = b_j, \quad j = 1, \cdots, p, \end{aligned} \tag{6.1.17}$$

其中 $f(x)$, $g_i(x), i = 1, \cdots, m$, 为 $\mathbf{R}^n$ 上的凸函数, $a_j, j = 1, \cdots, p$ 为 $\mathbf{R}^n$ 中向量, $b_j, j = 1, \cdots, p$ 为常数.

定理 6.1.7 (Fritz John 条件)　设 $x^* \in \mathbf{R}^n$ 为问题 (6.1.17) 的最优解, 则存在一组不全为零的常数 $\lambda_0, \lambda_1, \cdots, \lambda_m \geqslant 0$, $\mu_1, \cdots, \mu_p$, 使得

$$0 \in \lambda_0 \partial f(x^*) + \sum_{i=1}^{m} \lambda_i \partial g_i(x^*) + \sum_{j=1}^{p} \mu_j a_j, \tag{6.1.18}$$

$$\lambda_i g_i(x^*) = 0, \quad i = 1, \cdots, m. \tag{6.1.19}$$

证明　考虑下述辅助优化问题:

$$\begin{aligned} &\min F(x), \\ &\text{s.t. } a_j^{\mathrm{T}} x = b_j, \quad j = 1, \cdots, p, \end{aligned} \tag{6.1.20}$$

其中

$$F(x) = \max\{f(x) - f(x^*), g_1(x), \cdots, g_m(x)\}.$$

显然, 问题 (6.1.20) 是一个含有线性等式约束的凸规划问题, 由于 x^* 为问题 (6.1.17) 的最优解, 不难验证 x^* 也是问题 (6.1.20) 的最优解. 利用定理 6.1.4、定理 6.1.5 以及凸函数

$$F(x) = \max\{f(x) - f(x^*), g_1(x), \cdots, g_m(x)\}$$

次微分的表达式, 即得式 (6.1.18) 和式 (6.1.19). 定理得证.

6.2 Lipschitz 优化的最优性条件

所谓 Lipschitz 优化是指目标函数和约束条件均为局部 Lipschitz 函数的优化问题. 在非光滑优化最优性条件研究中, 由于使用了非光滑分析工具, 因而对不等式约束情形处理相对容易一些, 问题的难点则在等式约束部分.

6.2.1 不等式约束情形

考虑下述优化问题:

$$\begin{aligned} &\min f(x), \\ &\text{s.t. } g_i(x) \leqslant 0, \quad i = 1, \cdots, m, \end{aligned} \tag{6.2.1}$$

其中 $f(x), g_i(x), i = 1, \cdots, m$ 为 $\mathbf{R}^n$ 上的局部 Lipschitz 函数.

定理 6.2.1 (Fritz John 条件) 设 $x^* \in \mathbf{R}^n$ 为问题 (6.2.1) 的局部最优解, 则存在不全为零的常数 $\lambda_0, \lambda_1, \cdots, \lambda_m \geqslant 0$, 使得

$$0 \in \lambda_0 \partial f(x^*) + \sum_{i=1}^{m} \lambda_i \partial g_i(x^*), \tag{6.2.2}$$

$$\lambda_i g_i(x^*) = 0, \quad i = 1, \cdots, m. \tag{6.2.3}$$

证明 设

$$F(x) = \max\{f(x) - f(x^*), g_1(x), \cdots, g_m(x)\}.$$

显然, $F(x)$ 是 $\mathbf{R}^n$ 上的局部 Lipschitz 函数, 且在 x 的一个邻域内有

$$F(x) \geqslant 0, \quad F(x^*) = 0,$$

于是 x^* 是 $F(x)$ 的局部极小值点, 根据无约束问题广义梯度形式最优性条件得 $0 \in \partial F(x^*)$. 根据广义梯度运算规则, 有

$$\partial F(x^*) \subset \operatorname{co}\Big\{\partial f(x^*) \bigcup \Big(\bigcup_{i \in I(x^*)} \partial g_i(x^*)\Big)\Big\}, \tag{6.2.4}$$

其中

$$I(x^*) = \{i | g_i(x^*) = 0, i = 1, \cdots, m\},$$

由 $0 \in \partial F(x^*)$ 及式 (6.2.4) 得

$$0 \in \text{co}\Big\{\partial f(x^*) \bigcup \Big(\bigcup_{i \in I(x^*)} \partial g_i(x^*) \Big) \Big\}.$$

于是, 存在不全为零的常数 $\lambda_0, \lambda_i \geqslant 0, i \in I(x^*)$, 使得

$$0 \in \lambda_0 \partial f(x^*) + \sum_{i \in I(x^*)} \lambda_i \partial g_i(x^*), \quad i \in I(x^*). \tag{6.2.5}$$

令

$$\lambda_i = 0, \quad i \in \{1, \cdots, m\} \backslash I(x^*),$$

即得式 (6.2.2), 而此时式 (6.2.3) 也成立. 定理得证.

定理 6.2.1 为不等式约束 Lipschitz 优化广义梯度形式的 Fritz John 最优性必要条件, $\lambda_i, i = 1, \cdots, m$ 称为 Lagrange 乘子, 满足式 (6.2.2) 和式 (6.2.3) 的点称为 Fritz John 点. 当 $f(x)$, $g_i(x), i = 1, \cdots, m$ 为连续可微时, 此处的 Fritz John 条件即为通常非线性规划的 Fritz John 必要性条件.

6.2.2　等式与不等式约束情形

考虑下述同时含有等式与不等式约束的 Lipschitz 优化问题:

$$\begin{aligned} &\min f(x), \\ &\text{s.t.} g_i(x) \leqslant 0, \quad i = 1, \cdots, m, \\ &\quad h_j(x) = 0, \quad j = 1, \cdots, p, \end{aligned} \tag{6.2.6}$$

其中 $f(x)$, $g_i(x)$, $i = 1, \cdots, m$, $h_j(x), j = 1, \cdots, p$ 为 $\mathbf{R}^n$ 上的局部 Lipschitz 函数.

为建立优化问题 (6.2.6) 的最优性条件, 我们需要下面的 Ekeland 变分原理.

定理 6.2.2 (Ekeland 变分原理)　设 E 为完备的度量空间, $f(x)$ 为 E 上的非负下半连续函数, 给定 $x_0 \in \text{dom} f$, 其中 $\text{dom} f$ 为 $f(x)$ 定义域, $\varepsilon > 0$, 则存在 $\bar{x} \in E$, 使得

$$f(\bar{x}) + \varepsilon \|x_0 - \bar{x}\| \leqslant f(x_0), \tag{6.2.7}$$

$$f(\bar{x}) < f(x) + \varepsilon \|x - \bar{x}\|, \quad x \neq \bar{x}. \tag{6.2.8}$$

证明　不妨取 $\varepsilon = 1$. 定义下述 E 到 E 中子集的集值映射:

$$F(x) = \{y | f(y) + \|x - y\| \leqslant f(x)\}.$$

$f(x)$ 的下半连续性保证对每个 $x \in E, F(x)$ 是闭集, 而直接由 $F(x)$ 的定义有 $x \in F(x)$. 下面证明

$$y \in F(x) \Rightarrow F(y) \subset F(x). \tag{6.2.9}$$

如果 $x \notin \mathrm{dom} f$, 此时 $F(x) = E$, 故式 (6.2.9) 成立. 以下假设 $f(x)$ 取有限值. 选取 $y \in F(x)$ 和 $z \in F(y)$, 根据 $F(x)$ 的定义, 有

$$f(z) + \|y - z\| \leqslant f(y), \tag{6.2.10}$$

$$f(y) + \|x - y\| \leqslant f(x). \tag{6.2.11}$$

将式 (6.2.10) 和式 (6.2.11) 相加并利用三角不等式得

$$f(z) + \|x - z\| \leqslant f(x),$$

这说明 $z \in F(x)$, 故式 (6.2.9) 成立.

在 $\mathrm{dom} f$ 上定义函数 $v(y)$ 如下:

$$v(y) = \inf_{z \in F(y)} f(z), \tag{6.2.12}$$

易见下式成立:

$$\|x - y\| \leqslant f(x) - v(x),$$

这说明集合 $F(x)$ 的直径 $\mathrm{diam}(F(x))$ 满足

$$\mathrm{diam}(F(x)) \leqslant 2(f(x) - v(x)). \tag{6.2.13}$$

给定初始点 x_0, 定义点列 $\{x_n\}$ 使其满足

$$x_{n+1} \in F(x_n), \quad f(x_{n+1}) \leqslant v(x_n) + 2^{-n}.$$

根据 $F(x_{n+1}) \subset F(x_n)$ 及式 (6.2.9), 有

$$v(x_n) \leqslant v(x_{n+1}).$$

另一方面, 因为 $v(y) \leqslant f(y)$, 所以下述不等式成立

$$v(x_{n+1}) \leqslant f(x_{n+1}) \leqslant v(x_n) + 2^{-n} \leqslant v(x_{n+1}) + 2^{-n}, \tag{6.2.14}$$

进而有

$$0 \leqslant f(x_{n+1}) - v(x_{n+1}) \leqslant 2^{-n}, \tag{6.2.15}$$

式 (6.2.13) 和式 (6.2.15) 说明闭集 $F(x_n)$ 的半径收敛于 0. 再根据 $F(x_{n+1}) \subset F(x_n)$ 及空间 E 的完备性, 有

$$\bigcap_{n \geqslant 0} F(x_n) = \{\bar{x}\}.$$

因为 $\bar{x} \in F(x_0)$, 所以 $\bar{x}$ 满足式 (6.2.7). 另一方面,

$$\bar{x} \in F(x_n), \quad n \geqslant 0,$$

于是有

$$F(\bar{x}) \subset F(x_n),$$

进而有

$$F(\bar{x}) = \{\bar{x}\}.$$

这说明如果 $x \neq \bar{x}$, 则 $x \notin F(\bar{x})$, 于是

$$f(\bar{x}) < f(x) + \|\bar{x} - x\|,$$

式 (6.2.8) 成立. 定理得证.

Ekeland 变分原理是极值问题中的基本定理, 有多种等价形式.

推论 6.2.1　在定理 6.2.2 中条件成立下, 如果 $\varepsilon > 0, \lambda > 0$, x_0 满足 $f(x_0) \leqslant \inf f(x) + \varepsilon\lambda$, 则存在 $\bar{x} \in E$, 使得

$$f(\bar{x}) \leqslant f(x_0), \quad \|x_0 - \bar{x}\| \leqslant \lambda, \quad f(\bar{x}) \leqslant f(x) + \varepsilon\|x - \bar{x}\|.$$

下面给出具有等式与不等式约束的 Lipschitz 优化问题 Fritz John 条件.

定理 6.2.3 (Fritz John 条件)　设 $x^* \in \mathbf{R}^n$ 是问题 (6.2.6) 的最优解, 则存在一组不全为零的常数 $\lambda_i \geqslant 0$, $i = 0, 1, \cdots, m$, μ_j, $j = 1, \cdots, p$, 使得

$$0 \in \lambda_0 \partial f(x^*) + \sum_{i=1}^{m} \lambda_i \partial g_i(x^*) + \sum_{j=1}^{p} \mu_j \partial h_j(x^*), \tag{6.2.16}$$

$$\lambda_i g_i(x^*) = 0, \quad i = 1, \cdots, m. \tag{6.2.17}$$

证明　给定 $\varepsilon > 0$, 定义集合 T 和函数 $F(x)$ 如下:

$$T = \{t = (\lambda_0\ , \lambda, \mu)\, \mathbf{R}^{1+m+p} | \lambda_0, \lambda \geqslant 0, \|(\lambda_0\ \lambda\ \mu)\| = 1\},$$

$$F(x) = \max_{T}\{(\lambda_0\ , \lambda, \mu)^{\mathrm{T}} (f(x) - f(x^*) + \varepsilon, g(x), h(x))\}, \tag{6.2.18}$$

其中

$$\lambda \in \mathbf{R}^m, \quad \mu \in \mathbf{R}^p, \quad g(x) = (g_1(x), \cdots, g_m(x))^{\mathrm{T}}, \quad h(x) = (h_1(x), \cdots, h_p(x))^{\mathrm{T}}.$$

显然, $F(x)$ 在点 x^* 附近是 Lipschitz 的, 且有

$$F(x^*) = \varepsilon.$$

另外还有

$$F(x) > 0,$$

若不然, 则存在 $y \in \mathbf{R}^n$, 使得

$$F(y) \leqslant 0,$$

故

$$g(y) \leqslant 0, \quad h(y) \leqslant 0, \quad f(y) \leqslant f(x^*) - \varepsilon,$$

故 x^* 不是问题 (6.2.6) 最优解, 与假设矛盾. 于是有

$$F(x^*) \leqslant \inf F(x) + \varepsilon,$$

根据推论 6.2.1, 存在 $u \in B(x^*, \sqrt{\varepsilon})$, 使得对任意 $x \in \mathbf{R}^n$, 有

$$F(u) \leqslant F(x) + \sqrt{\varepsilon}\|x - u\|,$$

即 $x = u$ 是函数 $F(x) + \sqrt{\varepsilon}\|x - u\|$ 的极值小点. 根据广义梯度形式最优性条件及 $\partial\|x - u\| \subset B(0, 1)$, 则有

$$0 \in \partial F(u) + B(0, \sqrt{\varepsilon}).$$

为估计广义梯度 $\partial F(u)$, 我们首先证明集值映射

$$(t, x) \to \partial_x L(x, t)$$

是闭的 (上半连续), 其中

$$t = (\lambda_0, \lambda, \mu), \quad L(x, t) = \lambda_0 f(x) + \lambda g(x) + \mu h(x).$$

注意到对 $t_1, t_2 \in T$, 函数

$$y \to L(y, t_1) - L(y, t_2) = (t_1 - t_2)^{\mathrm{T}}(f(x), g(x), h(x))$$

是 Lipschitz 的, $L|t_1 - t_2|$ 为其 Lipschitz 常数, 其中 L 为 $f(x)$, $g(x)$, $h(x)$ 的 Lipschitz 常数, 于是有

$$\partial_x L(y, t_1) \subset \partial_x L(y, t_2) + L\|t_1 - t_2\|B(0, 1),$$

这说明映射 $\partial_x L(y,t)$ 是闭的. 由于 $F(u)>0$, 则存在唯一的 $t_u\in T$, 使得式 (6.2.18) 中的极大点在 t_u 处达到, 则有

$$0\in\partial_x L(u,t_u)+B(0,\sqrt{\varepsilon}). \tag{6.2.19}$$

注意到, 如果 $g_i(u)<0$, 必有 $r_i=0$, 令 $\varepsilon_i\to 0$, 相应地 $u_i\to x$, 而 $\{t_{u_i}\}$ 中存在子列收敛到 T 中元素. 于是, 由式 (6.2.19) 及集值映射 $(t,x)\to\partial_x L(x,t)$ 的闭性即得定理结论. 定理得证.

定理 6.2.3 为具有等式与不等式约束 Lipschitz 优化广义梯度形式的 Fritz John 最优性必要条件, $\lambda_i, i=1,\cdots,m$ 为 Lagrange 乘子, 满足式 (6.2.16) 和式 (6.2.17) 的点称为 Fritz John 点. 当然, 在一定约束品性下也可建立相应的 Kuhn-Tucker 条件. 当 $f(x)$, $g_i(x), i=1,\cdots,m$, $h_j(x), j=1,\cdots,p$ 为连续可微时, 此处的 Fritz John 条件即为通常非线性规划的 Fritz John 条件.

6.3 拟可微优化的最优性条件

拟可微优化是指目标函数和约束函数均为拟可微函数的优化问题. 本节将利用拟微分给出拟可微优化的最优性条件.

6.3.1 不等式约束情形

考虑下述优化问题:

$$\begin{aligned}&\min f_0(x),\\&\text{s.t. } f_i(x)\leqslant 0,\quad i=1,\cdots,m,\end{aligned} \tag{6.3.1}$$

其中 $f_i(x), i=0,\cdots,m$ 均为 $\mathbf{R}^n$ 上的拟可微函数.

定理 6.3.1 设 $x^*\in\mathbf{R}^n$ 为问题 (6.3.1) 的局部最优解, 则有

$$-\sum_{i\in\{0\}\cup I(x^*)}\bar{\partial}f_i(x^*)\subset \mathrm{co}\Big\{\underline{\partial}f_i(x^*)-\sum_{j\in\{0\}\cup\{I(x^*)\setminus\{i\}\}}\bar{\partial}f_j(x^*)|i\in\{0\}\bigcup I(x^*)\Big\}, \tag{6.3.2}$$

其中

$$I(x^*)=\{i|f_i(x^*)=0, i\neq 0\}.$$

证明 令

$$F(x)=\max\{f_0(x)-f_0(x^*), f_i(x)|i=1,\cdots,m\},$$

显然 $F(x)$ 也是拟可微函数, $F(x^*)=0$, 且在 x 的一个邻域内有

$$F(x)\geqslant 0,$$

于是

$$F(x) \geqslant F(x^*),$$

也就是说, x^* 是 $F(x)$ 的局部极小点. 根据拟可微函数极小点性质, 有

$$-\bar{\partial} F(x^*) \subset \underline{\partial} F(x^*).$$

利用拟微分计算公式直接推导知, $-\bar{\partial}F(x^*)$ 即为式 (6.3.2) 左端, $\underline{\partial}F(x^*)$ 为式 (6.3.2) 右端, 故式 (6.3.2) 成立. 命题得证.

式 (6.3.2) 是不等式约束优化拟微分形式的一阶必要条件, 但是它应该属于几何形式的条件, 因为在此条件中没有 Lagrange 乘子. 尽管如此, 式 (6.3.2) 确实是光滑优化问题 Fritz John 条件的推广, 事实上, 对于光滑问题, 如果取拟微分为

$$\underline{\partial} f_i(x) = \{\nabla f_i(x)\}, \quad \bar{\partial} f_i(x) = \{0\}, \quad i = 1, \cdots, m,$$

则式 (6.3.2) 即为非线性规划的 Fritz John 条件.

下述定理给出问题 (6.3.1) 的一个含有 Lagrange 乘子的最优性必要条件.

定理 6.3.2(Fritz John 条件) 设 $x^* \in \mathbf{R}^n$ 为问题 (6.3.1) 的一个局部最优解, 则对任意一组超微分 $v_i \in \bar{\partial} f_i(x), i = 0, 1, \cdots, m$, 存在一组不全为零依赖于 $v = (v_1, \cdots, v_m)$ 的常数 $\lambda_i(v) \geqslant 0, i = 0, 1, \cdots, m$, 使得

$$0 \in \sum_{i=0}^{m} \lambda_i(v)(v_i + \underline{\partial} f_i(x^*)), \tag{6.3.3}$$

$$\lambda_i(v) f_i(x^*) = 0, \quad i = 1, \cdots, m. \tag{6.3.4}$$

证明 记

$$I(x^*) = \{i | f_i(x^*) = 0, 1 \leqslant i \leqslant m\}.$$

根据假设, x^* 为问题 (6.3.1) 的最优解, 则下述不等式组无解

$$f_i'(x^*; y) < 0, \quad y \in \mathbf{R}^n, \quad i \in I(x^*). \tag{6.3.5}$$

事实上, 若不等式组 (6.3.5) 有解, 记 $\bar{y}$ 为它的一个解, 则 $\bar{y}$ 为问题 (6.3.1) 在 x^* 点处的可行下降方向, 这与 x^* 为问题 (6.3.1) 的最优解相矛盾. 根据拟微分定义易见, 对于 $v_i \in \bar{\partial} f_i(x)$ 有

$$\begin{aligned}\max_{u \in \underline{\partial} f_i(x^*)} u^{\mathrm{T}} y + v_i^{\mathrm{T}} y &\geqslant \max_{u \in \underline{\partial} f_i(x^*)} u^{\mathrm{T}} y + \min_{v \in \bar{\partial} f_i(x^*)} v^{\mathrm{T}} y \\ &= f'(x; y), \quad y \in \mathbf{R}^n, \quad i \in I(x^*),\end{aligned}$$

于是不等式组 (6.3.5) 无解意味着下述不等式组:

$$\max_{u\in\underline{\partial} f_i(x^*)} u^{\mathrm{T}}y+v_i^{\mathrm{T}}y<0, \quad y\in\mathbf{R}^n, \quad i\in\{0\}\bigcup I(x^*) \tag{6.3.6}$$

无解. 等价地, 下述优化问题:

$$\begin{aligned}&\min z,\\ &\text{s.t.} \max_{u\in\underline{\partial} f_i(x^*)+v_i} u^{\mathrm{T}}y-z\leqslant 0, \quad i\in\{0\}\bigcup I(x^*)\end{aligned} \tag{6.3.7}$$

的最优值为零. 问题 (6.3.7) 为 $\mathbf{R}^{n+1}$ 上的凸规划, 其中 $(y,z)\in\mathbf{R}^{n+1}$ 为变量. 由问题 (6.3.7) 的最优值为零, 则存在 $\bar{y}\in\mathbf{R}^n$, 使得 $(\bar{y},0)\in\mathbf{R}^{n+1}$ 为其最优解. 直接计算次微分得

$$\partial z|_{(y,z)=(\bar{y},0)}=(0,\cdots,0,1),$$

$$\partial(\max_{u\in\underline{\partial} f_i(x^*)+v_i} u^{\mathrm{T}}y-z)|_{(y,z)=(\bar{y},0)}=(\underline{\partial} f_i(x^*)+v_i,-1), \quad i\in\{0\}\bigcup I(x^*).$$

根据凸规划的最优性条件, 存在不全为零的 $\bar{\lambda}_0(v),\lambda_i(v)\geqslant 0, i\in\{0\}\bigcup I(x^*)$, 使得

$$\begin{aligned}0\in &\bar{\lambda}_0(v)\partial z|_{(y,z)=(\bar{y},0)}+\sum_{i\in\{0\}\cup I(x^*)}\lambda_i(v)\partial(\max_{u\in\underline{\partial} f_i(x^*)+v_i} u^{\mathrm{T}}y-z)|_{(y,z)=(\bar{y},0)}\\ &=\bar{\lambda}_0(v)(0,\cdots,0,1)+\sum_{i\in\{0\}\cup I(x^*)}\lambda_i(v)(\underline{\partial} f_i(x^*)+v_i,-1)\\ &=\Big(\sum_{i\in\{0\}\cup I(x^*)}\lambda_i(v)(\underline{\partial} f_i(x^*)+v_i),\bar{\lambda}_0(v)-\sum_{i\in\{0\}\cup I(x^*)}\lambda_i(v)\Big),\end{aligned}$$

所以有

$$0\in\sum_{i\in\{0\}\cup I(x^*)}^{m}\lambda_i(v)(\underline{\partial} f_i(x^*)+v_i), \tag{6.3.8}$$

$$\bar{\lambda}_0(v)-\sum_{i\in\{0\}\cup I(x^*)}\lambda_i(v)=0. \tag{6.3.9}$$

因为

$$\bar{\lambda}_0(v),\lambda_i(v)\geqslant 0, \quad i\in\{0\}\cup I(x^*)$$

不全为零, 式 (6.3.9) 意味着 $\bar{\lambda}_0(v)\neq 0$(否则,

$$\sum_{i\in\{0\}\cup I(x^*)}\lambda_i(v)=0,$$

进而

$$\lambda_i(v)=0, \quad i\in\{0\}\bigcup I(x^*),$$

这与 $\lambda_0(v), \lambda_i(v) \in \{0\}\bigcup I(x^*)$ 不全为零矛盾). 再由式 (6.3.9) 易见

$$\sum_{i\in\{0\}\cup I(x^*)} \lambda_i(v) \neq 0,$$

则

$$\lambda_i(v), i \in \{0\}\bigcup I(x^*)$$

不全为零. 令

$$\lambda_i(v) = 0, \quad i \in \{1, \cdots, m\}\backslash I(x^*),$$

则由式 (6.3.8) 即得到所证结论. 定理得证.

定理 6.3.2 给出的最优解条件较定理 6.3.1 所给出的更接近于 Fritz John 条件, 事实上可以证明两者是等价的.

6.3.2 具有等式与不等式约束情形

考虑下述含等式与不等式约束条件的优化问题:

$$\begin{aligned} &\min f_0(x), \\ &\text{s.t. } f_i(x) \leqslant 0, \quad i = 1, \cdots, m, \\ &\qquad h_j(x) = 0, \quad j = 1, \cdots, p, \end{aligned} \tag{6.3.10}$$

其中 $f_i(x), i = 1, \cdots, m$ 和 $h_j(x), j = 1, \cdots, p$ 是 $\mathbf{R}^n$ 上的拟可微函数.

给定 $x \in \mathbf{R}^n$, 引入下述两个锥:

$$\begin{aligned} \Gamma(x) = &\left\{ y \in \mathbf{R}^n | \lambda \geqslant 0, x_k \to x, x_k \neq x, h_j(x_k) = 0, 1 \leqslant j \leqslant p, \right. \\ &\left. \frac{x_k - x}{\|x_k - x\|} \to y_0, y = \lambda y_0 \right\}, \\ \gamma(x) = &\left\{ y \in \mathbf{R}^n | h_j'(x; y) = 0, j = 1, \cdots, p \right\}. \end{aligned}$$

引理 6.3.1 假设 $x^* \in \mathbf{R}^n$ 为问题 (6.3.10) 的最优解, $f_i(x), i = 1, \cdots, m$ 在 x^* 处是一致方向可微的, $\Gamma(x^*) = \gamma(x^*)$, 则 $y = 0$ 是下述优化问题的最优解:

$$\begin{aligned} &\min F'(x^*; y), \\ &\text{s.t. } h_j'(x^*; y) = 0, \quad j = 1, \cdots, p, \end{aligned} \tag{6.3.11}$$

其中

$$F(x) = \max\left\{ f_0(x) - f_0(x^*), f_i(x) | i \in I(x^*) \right\}, I(x^*) = \left\{ i | f_i(x^*) = 0, 1 \leqslant i \leqslant m \right\}.$$

证明 用反证法. 假设 $y=0$ 不是优化问题 (6.3.11) 的最优解, 于是存在 $\bar{y}\in\mathbf{R}^n$ 满足 $\|\bar{y}\|=1$, 使得

$$F'(x^*;\bar{y})<0,$$

$$h_j'(x^*;\bar{y})=0,\quad j=1,\cdots,p,$$

这说明 $\bar{y}\in\gamma(x^*)$, 由 $\gamma(x^*)=\Gamma(x^*)$, 则有 $\bar{y}\in\Gamma(x^*)$. 根据 $\Gamma(x^*)$ 的定义, 存在序列 $\{x_k\}$, 使得

$$x_k\to x^*,\quad h_j(x_k)=0,\quad j=1,\cdots,p,\quad y_k=\frac{x_k-x^*}{\|x_k-x^*\|}\to\bar{y}.$$

令

$$t_k=\|x_k-x^*\|,$$

则有

$$x_k=x^*+t_ky_k,$$

利用 $f_i(x),i=1,\cdots,m$ 在 x^* 的一致方向可微性, 直接推导得

$$\begin{aligned}\frac{f_i(x_k)-f_i(x^*)}{t_k}&=\frac{f_i(x^*+t_ky_k)-f_i(x^*)}{t_k}\\&\to f_i'(x^*;\bar{y})\\&\leqslant\max\left\{f_i'(x^*;\bar{y})|i\in\{0\}\bigcup I(x^*)\right\}\\&=F'(x^*;\bar{y})<0,\quad\forall i\in\{0\}\bigcup I(x^*),\end{aligned}$$

故 k 足够大时有

$$f_i(x_k)<f_i(x^*),\quad\forall i\in\{0\}\bigcup I(x^*).$$

另一方面

$$h_j(x_k)=0,\quad j=1,\cdots,p,$$

这与 x^* 为问题 (6.3.10) 的最优解矛盾. 引理得证.

下面利用 Demyanov 差给出含等式与不等式约束条件拟可微优化的一种 Fritz John 条件形式的最优性条件.

定理 6.3.3 (Fritz John 条件) 假设 $x^*\in\mathbf{R}^n$ 为问题 (6.3.10) 的最优解, $f_i(x),i=0,1,\cdots,m$ 和 $h_j(x),j=1,\cdots,p$ 在 x^* 处是一致方向可微的, $\Gamma(x^*)=\gamma(x^*)$, 则存在不全为零的常数 $\lambda_i\geqslant 0,i=0,1,\cdots,m,\mu_j,j=1,\cdots,p$, 使得下式成立

$$0\in\sum_{i=0}^{m}\lambda_i(\underline{\partial}f_i(x^*)\dot{-}(-\bar{\partial}f_i(x^*)))+\sum_{j=1}^{p}\mu_j(\underline{\partial}h_j(x^*)\dot{-}(-\bar{\partial}h_j(x^*))),\tag{6.3.12}$$

$$\lambda_if_i(x^*)=0,\quad i=1,\cdots,m.\tag{6.3.13}$$

证明 根据引理 6.3.1, x^* 为问题 (6.3.10) 最优解, 则 $y=0$ 是问题 (6.3.11) 的最优解. 由于

$$F'(x^*;y)=\max\{f_i'(x^*;\bar{y})|i\in\{0\}\bigcup I(x^*)\},$$

易见, $(y,z)=0$ 是下述问题最优解:

$$\begin{aligned}&\min z\\&\text{s.t. } f_i'(x^*;z)\leqslant 0,\quad i\in\{0\}\bigcup I(x^*),\\&\qquad h_j'(x^*;y)=0,\quad j=1,\cdots,p,\end{aligned}\tag{6.3.14}$$

其中 $(y,z)\in\mathbf{R}^{n+1}$ 是变量. 拟可微函数的方向导数关于方向是两个凸函数的差, 因此是局部 Lipschitz 的, 所以优化问题 (6.3.14) 中涉及的函数都是局部 Lipschitz 函数. 根据 Lipschitz 优化中广义梯度形式的 Fritz John 条件, 存在一组不全为零的常数

$$\bar{\lambda}_0,\lambda_i\geqslant 0,\quad i\in\{0\}\bigcup I(x^*),\quad \mu_j,\quad j=1,\cdots,p,$$

使得

$$0\in\bar{\lambda}_0\partial z|_{(y,z)=0}+\sum_{i\in\{0\}\cup I(x^*)}\lambda_i\partial(f_i'(x^*;y)-z)|_{(y,z)=0}+\sum_{j=1}^{p}\mu_j\partial h_j'(x^*;y)|_{(y,z)=0}.\tag{6.3.15}$$

根据 Demyanov 差的性质, 由式 (6.3.15) 可得

$$\begin{aligned}0\in\bar{\lambda}_0(0,1)^{\mathrm{T}}+&\sum_{i\in\{0\}\cup I(x^*)}^{m}\lambda_i(\underline{\partial}f_i(x^*)\dot{-}(-\bar{\partial}f_i(x^*)),-1)\\&+\sum_{j=1}^{p}\mu_j(\underline{\partial}h_j(x^*)\dot{-}(-\bar{\partial}h_j(x^*)),0),\end{aligned}$$

进而有

$$0\in\sum_{i\in\{0\}\cup I(x^*)}^{m}\lambda_i(\underline{\partial}f_i(x^*)\dot{-}(-\bar{\partial}f_i(x^*)))+\sum_{j=1}^{p}\mu_j(\underline{\partial}h_j(x^*)\dot{-}(-\bar{\partial}h_j(x^*))),\tag{6.3.16}$$

$$\bar{\lambda}_0=\sum_{i\in\{0\}\cup I(x^*)}\lambda_i.\tag{6.3.17}$$

式 (6.3.17) 说明

$$\lambda_i,\quad i\in\{0\}\bigcup I(x^*),\quad \mu_j,\quad j=1,\cdots,p$$

不全为零 (否则式 (6.3.17) 意味着 $\bar{\lambda}_0=0$), 取

$$\lambda_i=0,\quad \forall i\in\{1,\cdots,m\}\setminus I(x^*),$$

这样就得到式 (6.3.12) 和式 (6.3.13). 定理得证.

如果只含不等式约束条件的拟可微优化问题, 则在其 Fritz John 条件建立中, 相关函数的一致方向可微性条件可以去掉.

定理 6.3.4 设 $x^* \in \mathbf{R}^n$ 为不等式约束问题 (6.3.1) 的局部最优解, 则存在不全为零的常数 $\lambda_i \geqslant 0, i=0,1,\cdots,m$, 使得下式成立

$$0 \in \sum_{i=1}^{m} \lambda_i(\underline{\partial} f_i(x^*) - (-\bar{\partial} f_i(x^*))), \tag{6.3.18}$$

$$\lambda_i f_i(x^*) = 0, \quad i=1,\cdots,m. \tag{6.3.19}$$

证明 x^* 为问题 (6.3.1) 的最优解, 按定理 6.3.1 的证明有下述不等式:

$$f_i'(x^*;y) < 0, \quad i \in \{0\} \bigcup I(x^*)$$

无解, 进而下述优化问题:

$$\begin{aligned} &\min z, \\ &\text{s.t. } f_i'(x^*;y) - z \leqslant 0, \quad i \in \{0\} \bigcup I(x^*) \end{aligned}$$

最优值为零, 类似于定理 6.3.3 的证明即可得到式 (6.3.18) 和式 (6.3.19). 定理得证.

利用 Demyanov 差给出的最优性条件不但含有 Lagrange 乘子, 而且乘子也不再依赖于具体的超微分, 这是目前拟可微优化中最接近光滑优化 Fritz John 条件形式的一种最优性条件.

第 7 章　非光滑优化算法

所谓非光滑优化算法, 是指基于某种广义微分 (例如凸函数次微分、局部 Lipschitz 函数的广义梯度等) 理论所构造、设计的求解非光滑优化的算法, 而不包括非线性规划中的直接法 (例如单纯形法、坐标轮换法、随机搜索法等). 到目前为止, 针对一般形式非光滑优化的算法还仅仅是 "概念性算法" 而非 "可执行算法", 这是因为在非光滑优化算法中要求在每一迭代点处能够计算到所涉及函数次微分中的一个元素, 以确定迭代 (下降) 方向, 对于一般的非光滑函数, 这一点是无法做到的. 本章介绍非光滑优化的算法, 我们仅限于无约束问题.

7.1　下 降 方 向

7.1.1　局部 Lipschitz 函数的下降方向

考虑无约束优化问题

$$\min_{x\in\mathbf{R}^n} f(x), \tag{7.1.1}$$

其中 $f(x)$ 是 $\mathbf{R}^n$ 上的实函数. 经典的求解无约束优化问题 (7.1.1) 算法的迭代公式如下:

$$x_{k+1} = x_k + \lambda_k d_k, \tag{7.1.2}$$

其中 $d_k \in \mathbf{R}^n$ 是搜索方向, 通常要求它是目标函数的下降方向; λ_k 是步长, 通常要求满足某种搜索准则, 例如, 在精确搜索中 λ_k 要满足:

$$f(x_k + \lambda_k d_k) = \min_{\lambda\geqslant 0} f(x_k + \lambda d_k). \tag{7.1.3}$$

迭代法 (7.1.2) 产生的点列期望收敛于问题 (7.1.1) 的最优解, 至少应该是它的一个稳定点.

在上述算法框架中, 一个关键的问题是如何选取搜索方向 d_k. 当 $f(x)$ 是连续可微函数时, d_k 可选取为负梯度方向, 即

$$d_k = -\nabla f(x_k),$$

这就是通常的最速下降方向, 如果此时一维搜索按式 (7.1.3) 给出的规则, 则相应的迭代法为最速下降法. 然而, 对于非光滑情形, 如果 ξ 是某种次微分的一个元素, 一

般来讲方向 $d=-\xi$ 并不是下降方向, 更不用说是最速下降方向了. 另一方面, 即使能够找到最速下降方向, 相应的非光滑优化最速下降法也无法保证收敛性. 然而, 下述定理给出了一个利用广义梯度确定局部 Lipschitz 函数下降的方法.

定理 7.1.1　设 $f(x)$ 是 $\mathbf{R}^n$ 上的局部 Lipschitz 函数, 如果 $0 \notin \partial f(x)$, 则存在 $\bar{\xi} \in \partial f(x)$ 满足:

$$\|\bar{\xi}\| = \min_{\xi \in \partial f(x)} \|\xi\|, \tag{7.1.4}$$

使得 $d=-\bar{\xi}$ 是 $f(x)$ 在 x 点的一个下降方向下.

证明　由于广义梯度 $\partial f(x)$ 是紧集, 故一定存在 $\bar{\xi} \in \partial f(x)$ 满足式 (7.1.4). 又 $\partial f(x)$ 是凸集, 因此对任意 $\xi \in \partial f(x)$ 和 $\lambda \in (0,1)$, 有

$$\bar{\xi} + \lambda(\xi - \bar{\xi}) \in \partial f(x).$$

进一步, 由式 (7.1.4) 得

$$\|\bar{\xi} + \lambda(\xi - \bar{\xi})\| = \|\bar{\xi}\|^2 + 2\lambda(\xi - \bar{\xi})^{\mathrm{T}}\bar{\xi} + \lambda^2\|\xi - \bar{\xi}\|^2 \geqslant \|\bar{\xi}\|^2,$$

于是

$$(\xi - \bar{\xi})^{\mathrm{T}}\bar{\xi} \geqslant -\frac{\lambda}{2}\|\xi - \bar{\xi}\|^2, \quad \forall \xi \in \partial f(x).$$

在上式中令 $\lambda \to 0^+$, 得

$$\|\bar{\xi}\|^2 \leqslant \xi^{\mathrm{T}}\bar{\xi}, \quad \forall \xi \in \partial f(x). \tag{7.1.5}$$

根据中值定理, 存在 $x' \in (x; x - t\bar{\xi})$, 使得

$$f(x - t\bar{\xi}) - f(x) \in \partial f(x')^{\mathrm{T}}(-t\bar{\xi}). \tag{7.1.6}$$

根据广义梯度的上半连续性, 对充分小的 t 有

$$\partial f(x') \in \partial f(x) + \frac{1}{2}\|\bar{\xi}\|B(0,1),$$

于是由式 (7.1.5) 和式 (7.1.6), 对任意 $\xi_1 \in \partial f(x')$ 有

$$\xi_1^{\mathrm{T}} d \leqslant -\|\bar{\xi}\| + \frac{1}{2}\|\xi\|^2 = \frac{1}{2}\|\bar{\xi}\|^2. \tag{7.1.7}$$

结合式 (7.1.6) 和式 (7.1.7), 对充分小的 t 有

$$f(x - t\bar{\xi}) - f(x) \leqslant -\frac{1}{2}\|\xi\|^2,$$

这说明 $-\bar{\xi}$ 是 $f(x)$ 在点 x 处的下降方向. 定理得证.

定理 7.1.1 说明, 广义梯度集到原点最近点的负方向是一个下降方向, 一般情况下它不是最速下降方向. 有时也称其为广义最速下降方向. 基于定理 7.1.1, 我们可以给出求解局部 Lipschitz 函数 $f(x)$ 极小化的广义最速下降方法.

算法 7.1.1

步 1. 给定初始点 $x_0 \in \mathbf{R}^n$, 令 $k=0$;

步 2. 求 $\xi_k^* \in \partial f(x_k)$, 满足 $\|\xi_k^*\| = \min\limits_{\xi \in \partial f(x_k)} \|\xi\|$, 若 $\xi_k^*=0$, 则停止; 否则转步 3;

步 3. 令 $d_k = -\xi_k^*$, 求步长 λ_k 满足 $f(x_k+\lambda_k d_k) = \min\limits_{\lambda \geqslant 0} f(x_k+\lambda d_k)$;

步 4. 令 $x_{k+1} = x_k + \lambda_k d_k$, $k=k+1$, 转步 2.

尽管上述算法确实是一个下降算法, 但是收敛性还是无法保证. 当然, 如果 $f(x)$ 为连续可微时, 算法 7.1.1 即为经典的最速下降法.

7.1.2 凸函数下降方向的计算

设 $f(x)$ 为 $\mathbf{R}^n$ 上的凸函数, d 为 $f(x)$ 在点 x 的下降方向, 则存在足够小的 $t>0$, 使得

$$f(x+td) < f(x),$$

而另一方面根据次微分定义, 有

$$f(x+td) \geqslant f(x) + t\xi^{\mathrm{T}} d, \quad \forall \xi \in \partial f(x), \tag{7.1.8}$$

进而有

$$f(x+td) \geqslant f(x) + tf'(x;d). \tag{7.1.9}$$

结合式 (7.1.8), 式 (7.1.9) 和关系式

$$f(x+td) < f(x)$$

知, $d \in \mathbf{R}^n$ 为凸函数 $f(x)$ 在点 x 的下降方向, 等价于

$$f'(x;d) < 0,$$

也等价于

$$\xi^{\mathrm{T}} d < 0, \quad \forall \xi \in \partial f(x). \tag{7.1.10}$$

下面基于式 (7.1.10) 对下降方向的刻画计算凸函数 $f(x)$ 在非极小点 x 的一个下降方向. 假设已经得到 $f(x)$ 在点 x 的 k 个次微分, 即

$$\xi_1, \cdots, \xi_k \in \partial f(x),$$

记

$$S_k = \mathrm{co}\{\xi_1, \cdots, \xi_k\},$$

显然 $S_k \subset \partial f(x)$. 若 d 为 $f(x)$ 在 x 点的下降方向, 则必有

$$
\begin{aligned}
0 > f'(x;d) &= \max_{\xi \in \partial f(x)} \xi^{\mathrm{T}} d \\
&\geqslant \max_{\xi \in S_k} \xi^{\mathrm{T}} d \geqslant \max_{1 \leqslant i \leqslant k} \xi_i^{\mathrm{T}} d,
\end{aligned}
$$

因此, 下降方向 d 必需满足

$$
\xi_i^{\mathrm{T}} d < 0, \quad i = 1, \cdots, k. \tag{7.1.11}
$$

首先求解线性不等式组 (7.1.11), 记 d 为它的解, 通过线搜索检验 d 是否为下降方向. 如果 d 不是下降方向, 则必有 $t > 0$, 使得

$$
f(x) \geqslant f(x + td).
$$

根据凸函数次微分的定义, 对于 $\xi_t \in \partial f(x + td)$ 有

$$
\begin{aligned}
f(x) &\geqslant f(x + td) + \xi_t^{\mathrm{T}}(x - x - td) \\
&= f(x + td) - t\xi_t^{\mathrm{T}} d,
\end{aligned}
$$

于是 $\xi_t^{\mathrm{T}} d \geqslant 0$. 由于 $\partial f(x + td)$ 有界, $\{\xi_t\}$ 必有一个收敛子列, 不妨记为 $\{\xi_t\}$ 本身, 记

$$
\xi_t \to \xi^*, \quad t \to 0^+,
$$

根据次微分的上半连续性必有

$$
\xi^* \in \partial f(x) \quad (\text{因为 } x + td \to x).
$$

由于

$$
\xi_t^{\mathrm{T}} d \geqslant 0,
$$

于是

$$
\xi^* d \geqslant 0.
$$

至此, 我们找到了 $\partial f(x)$ 的一个元素 ξ^*, 它不满足式 (7.1.11), 记

$$
\xi_{k+1} = \xi^*,
$$

将其加到 S_k 中得

$$
S_{k+1} = \mathrm{co}\{\xi_1, \cdots, \xi_{k+1}\},
$$

然后重复这一过程, 期望得到函数 $f(x)$ 在点 x 的一个下降方向.

下面给出计算 $f(x)$ 在点 x 下降方向算法的具体步骤.

算法 7.1.2(计算凸函数下降方向)

步 0. 给定 $x \in \mathbf{R}^n$, 计算 $f(x)$ 和 $\xi_1 \in \partial f(x)$, 令 $k=1$;

步 1. 令 $S_k = \text{co}\{\xi_1, \cdots, \xi_k\}$, 求解下述优化问题:

$$\min\{\|d\|^2 | -d \in S_k\}, \tag{7.1.12}$$

记 d_k 为它的解, 如果 $d_k = 0$, x 是 $f(x)$ 的极小点, 无下降方向, 停止; 否则转下一步;

步 2. 对方向 d_k 进行线搜索, 如果有 $t > 0$, 使得 $f(x+td_k) < f(x)$, 停止, d_k 是下降方向, 否则转下一步;

步 3. 令 $t \to 0^+$, 确定 $\xi_{k+1} \in \partial f(x)$, 使得 $\xi_{k+1}^{\mathrm{T}} d_k \geqslant 0$, 令 $k = k+1$, 转步 1.

根据前面的讨论知, 步 3 中的 ξ_k 是可以得到的. 另一方面, (7.1.12) 是一个二次规划问题, 因此, 如果在一点处可以计算到 $f(x)$ 的一个次微分, 算法 7.1.2 是可实现的. 下述定理说明利用算法 7.1.2 可在有限步计算到 $f(x)$ 的一个次微分.

定理 7.1.2 如果 x 不是 $f(x)$ 的极小值点, 则在算法 7.1.2 中必存在有限数 k, 使得

$$f'(x; d_k) < 0.$$

证明 我们证明这样一个事实: 设 $\{x_k\}$ 和 $\{y_k\}$ 为 $\mathbf{R}^n$ 中两个点列, 满足

$$(x_j - x_{k+1})^{\mathrm{T}} y_k \geqslant \|y_k\|^2, \quad k \geqslant 1, \quad j = 1, \cdots, k, \tag{7.1.13}$$

如果 $\{x_k\}$ 有界, 则

$$y_k \to 0, \quad k \to \infty.$$

对式 (7.1.13) 应用 Cauchy-Schwarz 不等式得

$$\|y_k\| \leqslant \|x_j - x_{k+1}\|, \quad k \geqslant 1, \quad j = 1, \cdots, k.$$

如果 $y_k \to 0$ 不成立, 则存在常数 $\delta > 0$ 和 $\{y_k\}$ 中子列, 不妨记为 $\{y_k\}$ 本身, 使得

$$0 < \delta \leqslant \|y_k\| \leqslant \|x_j - x_{k+1}\|.$$

由 $\|x_k\|$ 的有界性, 则存在收敛子列, 在上式中再选取 $\{x_k\}$ 的收敛子列, 于是上式右端项趋于零, 得到矛盾, 故 $y_k \to 0$.

在算法 7.1.2 中, $-d_k$ 为 0 到 S_k 上的投影, 记 $-d_k = \widehat{\xi_k}$, 根据第 1 章点到凸集投影性质, 有

$$\xi^{\mathrm{T}} \widehat{\xi_k} \geqslant \|\widehat{\xi_k}\|^2, \quad \xi \in S_k. \tag{7.1.14}$$

再根据算法 7.1.2 中步 3, 得

$$\xi_{k+1}^{\mathrm{T}} \widehat{\xi_k} \leqslant 0. \tag{7.1.15}$$

在式 (7.1.14) 中, 选取

$$\xi = \xi_j, \quad j = 1, \cdots, k,$$

于是有

$$\xi_j^{\mathrm{T}} \widehat{\xi_k} \geqslant \|\widehat{\xi_k}\|^2, \quad j = 1, \cdots, k, \tag{7.1.16}$$

结合式 (7.1.5) 和式 (7.1.6), 得到

$$(\xi_j - \xi_{k+1})^{\mathrm{T}} \widehat{\xi_k} \geqslant \|\widehat{\xi_k}\|^2, \quad j = 1, \cdots, k.$$

在式 (7.1.13) 中, 令

$$x_k = \xi_k, y_k = \widehat{\xi_k},$$

注意到 $\{\xi_k\}$ 的有界性, 利用前面得到的结论有

$$\widehat{\xi_k} \to 0.$$

再由 $\widehat{\xi_k} \in \partial f(x)$ 及集值映射 $\partial f(x)$ 的上半连续性, 则有

$$0 \in \partial f(x),$$

与 x 不是 $f(x)$ 的极小值点. 定理得证.

相对来讲, 算法 7.1.2 给出的方法, 要较广义最速下降方向容易实现, 因为它不需要知道整个次微分集合, 而只需在每一点处计算到一个次微分即可.

7.2　凸规划的次梯度法

本节在 $f(x)$ 为凸函数时给出求解无约束优化 (7.1.1) 的一个算法, 称为次梯度法, 它是非光滑优化最早和最基本的算法之一.

7.2.1　算法步骤

首先给出解问题 (7.1.1) 算法的具体步骤.

算法 7.2.1(次梯度法)

步 0. 选取数列 $\{\lambda_k\}$, 满足 $\lambda_k > 0$, $\lambda_k \to 0 (k \to \infty)$, $\sum\limits_{k=0}^{\infty} \lambda_k = +\infty$, 给定 $x_0 \in \mathbf{R}^n$, 令 $k = 0$;

步 1. 计算 $\xi_k \in \partial f(x_k)$, 若 $\xi_k = 0$, 则停止; 否则转步 2;

步 2. 令 $x_{k+1} = x_k - \dfrac{\lambda_k}{\|\xi_k\|} \xi_k$, $k = k + 1$, 转步 1.

从算法步骤可以看出, 次梯度法迭代简单, 比较容易实现, 在每一迭代点处只要能够计算到目标函数次微分中一个元素就可以. 另一方面, 次梯度法不需要线搜索, 步长 λ_k 事先给出, 只要满足条件

$$\lambda_k > 0, \quad \lambda_k \to 0, \quad k \to \infty, \quad \sum_{k=0}^{\infty} \lambda_k = +\infty$$

即可. 在这里, 条件

$$\lambda_k \to 0, \quad k \to \infty$$

是要使点列 $\{x_k\}$ 收敛所必需的, 条件

$$\sum_{k=0}^{\infty} \lambda_k = +\infty$$

是要使点列 $\{x_k\}$ 全局收敛所必需的. 次梯度法甚至不是下降算法, 因为 $-\frac{\xi_k}{\|\xi_k\|}$ 并不一定是凸函数 $f(x)$ 的下降方向.

作为最早提出的非光滑优化算法, 次梯度法有许多不足之处, 一是收敛速度慢, 二是没有给出停止准则.

7.2.2 收敛性分析

首先给出一个收敛性中有用的命题.

命题 7.2.1 假设 $\{x_k\}$ 为算法 7.2.1 产生的点列, 如果

$$x_k \notin S^* = \{x \in \mathbf{R}^n | f(x) = \min_{y \in \mathbf{R}^n} f(y)\},$$

则对任意 $x^* \in S^*$, $\xi_k \in \partial f(x_k)$, 必存在常数 $T_k > 0$, 使得

$$\left\| x_k - \frac{\lambda}{\|\xi_k\|}\xi_k - x^* \right\| < \|x_k - x^*\|, \quad \forall \lambda \in (0, T_k). \tag{7.2.1}$$

证明 直接计算得

$$\left\| x_k - \frac{\lambda}{\|\xi_k\|}\xi_k - x^* \right\|^2 = \|x_k - x^*\|^2 + \frac{2\lambda}{\|\xi_k\|}\xi_k^{\mathrm{T}}(x^* - x_k) + \lambda^2.$$

若

$$\lambda \leqslant -\frac{2}{\|\xi_k\|}\xi_k^{\mathrm{T}}(x^* - x_k),$$

则式 (7.2.1) 成立. 注意到

$$\xi_k \in \partial f(x_k), \quad x_k \notin S^*,$$

根据凸函数次微分定义得

$$\xi_k^{\mathrm{T}}(x^* - x_k) \leqslant f(x^*) - f(x_k) < 0.$$

令

$$T_k = -\frac{2}{\|\xi_k\|}\xi_k^{\mathrm{T}}(x_k - x^*),$$

易见 $T_k > 0$, 此时式 (7.2.1) 成立. 命题得证.

下述定理给初次梯度法的一个收敛性结果.

定理 7.2.1　设问题 (7.1.1) 的解集

$$S^* = \{x \in \mathbf{R}^n | f(x) = \min_{y\in\mathbf{R}^n} f(y)\}$$

非空且有界, 则算法 (7.2.1) 产生的点列 $\{x_k\}$ 有下述性质:

$$\lim_{k\to\infty} d_{S^*}(x_k) = 0, \tag{7.2.2}$$

其中 $d_{S^*}(\cdot)$ 为距离函数.

证明　记

$$f^* = \min_{x\in\mathbf{R}^n} f(x),$$

根据 $f(x)$ 的凸性, 必存在连续函数 $\delta(\varepsilon)$, 使得

$$f(x_k) \leqslant f^* + \varepsilon \tag{7.2.3}$$

对任何满足

$$d_{S^*}(x_k) \leqslant \delta(\varepsilon), \tag{7.2.4}$$

的 x_k 成立. 对每个 k, 定义

$$\varepsilon_k = f(x_k) - f^* \geqslant 0. \tag{7.2.5}$$

如果 $\varepsilon_k > 0$, 则有

$$\begin{aligned}
\|x_{k+1} - x^*\|^2 &= \|x_k - x^*\|^2 + \lambda_k^2 - 2\lambda_k (x_k - x^*)^{\mathrm{T}} \frac{\xi_k}{\|\xi_k\|} \\
&= \|x_k - x^*\|^2 + \lambda_k^2 - 2\delta(\varepsilon_k)\lambda_k \\
&\quad -2\lambda_k \left(x_k - x^* - \delta(\varepsilon_k)\frac{\xi_k}{\|\xi_k\|}\right)^{\mathrm{T}} \frac{\xi_k}{\|\xi_k\|} \\
&\leqslant \|x_k - x^*\|^2 + \lambda_k^2 - 2\delta(\varepsilon_k)\lambda_k,
\end{aligned} \tag{7.2.6}$$

故

$$(d_{S^*}(x_{k+1}))^2 - (d_{S^*}(x_k))^2 \leqslant -\lambda_k(2\delta(\varepsilon_k) - \lambda_k). \tag{7.2.7}$$

定义 $\delta(0)=0$, 则上式对于一切 k 都成立. 对式 (7.2.7) 两边求和得

$$\lim_{k\to\infty}\inf\delta(\varepsilon_k)=0, \tag{7.2.8}$$

所以

$$\lim_{k\to\infty}\inf d_{S_k}(x_k)=0. \tag{7.2.9}$$

如果式 (7.2.2) 不成立, 则必存在正常数 $\delta'>0$ 和无穷多个 k, 使得

$$d_{S_k}(x_{k+1})>d_{S_k}(x_k), \tag{7.2.10}$$

$$\varepsilon_k>\delta'. \tag{7.2.11}$$

由式 (7.2.10) 和式 (7.2.7) 知

$$2\delta(\varepsilon_k)<\lambda_k \tag{7.2.12}$$

对于充分大的 k 成立. 式 (7.2.11) 与式 (7.2.12) 相矛盾, 这说明式 (7.2.2) 成立. 定理得证.

7.3 凸规划的割平面法

本节在 $f(x)$ 为凸函数时给出求解无约束优化 (7.1.1) 的割平面法算法, 它也是非光滑优化较基本的算法之一.

对于凸函数 $f(x)$, 有

$$f(x)=\max_{y\in\mathbf{R}^n}\max_{\xi\in\partial f(y)}(f(y)+\xi^{\mathrm{T}}(x-y)).$$

所以, 求 $f(x)$ 的极小点就等价于求解问题:

$$\begin{aligned}&\min\ v,\\&\text{s.t. } f(y)+\xi^{\mathrm{T}}(x-y)\leqslant v,\quad \forall\xi\in\partial f(y),\quad y\in\mathbf{R}^n.\end{aligned} \tag{7.3.1}$$

假设 $x_i,\ i=1,\cdots,k$ 是已得到的迭代点, 在第 k 次迭代中求解问题:

$$\begin{aligned}&\min\ v,\\&\text{s.t. } f(x_i)+\xi_i^{\mathrm{T}}(x-x_i)\leqslant v,\quad i=1,\cdots,k.\end{aligned} \tag{7.3.2}$$

显然, 问题 (7.3.1) 是一个线性规划, 且是问题 (7.1.1) 最优解得的一个近似, 重复这个过程就是极小化凸函数的割平面法.

下面给出算法的具体步骤.

算法 7.3.1(割平面法)

步 1. 给定初始点 $x_1 \in S$, S 是一给定的凸多面体, 令 $k=1$;

步 2. 求 $\xi_k \in \partial f(x_k)$, 若 $\xi_k = 0$, 则停止; 否则, 转步 3;

步 3. 在 S 上求解问题 (7.3.1), 即

$$
\begin{aligned}
&\min\ v, \\
&\text{s.t.}\ f(x_i) + \xi_i^{\mathrm{T}}(x - x_i) \leqslant v, \quad i = 1, \cdots, k, \\
&\quad\ x \in S,
\end{aligned}
$$

得最优解 (x_{k+1}, v_{k+1}), 令 $k=k+1$, 转步 2.

在算法 7.3.1 中, 每一次迭代增加一个约束, 从几何上看, 是用一个超平面将 S 中不包含解的部分割掉. 这在算法具体实现上将无限地增加约束, 因此计算量很大, 但它的收敛性却是可以保证的.

第 8 章　非光滑方程组及非线性互补问题

求解非光滑方程组, 特别是它的牛顿法, 是 20 世纪 90 年代初发展起来的. 这种方法产生的原因一方面是由于非光滑分析理论不断完善, 另一方面是由于它可有效地求解互补以及变分不等式问题. 本章将介绍求解非光滑方程组的牛顿法及在非线性互补中的应用.

8.1　半光滑函数及性质

对于一般的 Lipschitz 函数, 牛顿法的收敛性无法保证, 为此引入 Lipschitz 函数类中的一个子类 —— 半光滑函数.

定义 8.1.1　设 $H(x)$ 为 $\mathbf{R}^n$ 到 $\mathbf{R}^m$ 上的局部 Lipschitz 函数, 若对任意 $d \in \mathbf{R}^n$, 下述极限

$$\lim_{\substack{V \in \partial H(x+td') \\ d' \to d, t \to 0^+}} Vd' \tag{8.1.1}$$

总存在, 其中 $\partial H(x)$ 为 $H(x)$ 的广义 Jacobi, 则称 $H(x)$ 在 x 点是半光滑的.

定理 8.1.1　设 $H(x)$ 为 $\mathbf{R}^n$ 到 $\mathbf{R}^m$ 上的局部 Lipschitz 函数, 如果极限

$$\lim_{\substack{V \in \partial H(x+d) \\ t \to 0^+}} Vd \tag{8.1.2}$$

总存在, 则 $H(x)$ 的经典方向导数存在, 且有

$$H'(x;d) = \lim_{t \to 0^+} \frac{H(x+td) - H(x)}{t} = \lim_{\substack{V \in \partial H(x+td) \\ t \to 0^+}} Vd. \tag{8.1.3}$$

证明　$H(x)$ 的 Lipschitz 性质保证差商

$$\frac{H(x+td) - H(x)}{t}$$

有界, 因此它必有极限点, 设 l 是它的一个极限点, 则存在数列 $\{t_i\}$, 使得

$$l = \lim_{t_i \to 0^+} \frac{H(x+t_i d) - H(x)}{t_i}.$$

下面证明 l 等于式 (8.1.2). 根据广义 Jacobi 形式的中值定理知

$$\frac{H(x+t_i d) - H(x)}{t_i} \in \text{co}\partial H([x, x+t_i d])d.$$

对每个 i, 由 Caratheodory 定理 (定理 1.1.3), 存在

$$t_i^k \in [0, t_i], \quad \lambda_i^k \geqslant 0, \quad V_i^k \in \partial H([x, x + t_i^k d]), \quad k = 0, 1, \cdots, m,$$

满足

$$\sum_{k=0}^{m} \lambda_i^k = 1,$$

使得

$$\frac{H(x + t_i d) - H(x)}{t_i} = \sum_{k=0}^{m} \lambda_i^k V_i^k d.$$

序列 $\{t_i^k\}$ 和 $\{\lambda_i^k\}$ 有界, 则在指标集 $\{i\}$ 中存在子列, 不妨记为 $\{i\}$ 本身, 使得

$$t_i^k \to 0^+, \quad \lambda_i^k \to \lambda^k, \quad i \to +\infty, \quad k = 0, 1, \cdots, m.$$

显然

$$\lambda^k \geqslant 0, \quad k = 0, 1, \cdots, m,$$

且有

$$\sum_{k=0}^{m} \lambda_i^k = 1.$$

直接推导得

$$\begin{aligned}
l &= \lim_{i \to \infty} \frac{H(x + t_i d) - H(x)}{t_i} = \lim_{i \to \infty} \sum_{k=0}^{m} \lambda_i^k V_i^k d \\
&= \sum_{k=0}^{m} \lim_{i \to \infty} \lambda_i^k \lim_{i \to \infty} V_i^k d = \left(\sum_{k=0}^{m} \lambda^k \right) \lim_{\substack{V \in \partial H(x + td) \\ t \to 0^+}} V d \\
&= \lim_{\substack{V \in \partial H(x + td) \\ t \to 0^+}} V d,
\end{aligned}$$

这说明式 (8.1.3) 成立. 定理得证.

引理 8.1.1　设 $H(x)$ 为 $\mathbf{R}^n$ 到 $\mathbf{R}^m$ 上的局部 Lipschitz 函数, 如果 $H(x)$ 在 x 点是半光滑的, 则对任意 $d \in \mathbf{R}^n$, 存在 $V \in \partial H(x)$, 使得 $H'(x; d) = Vd$.

证明　由于 $H(x)$ 在 x 点是半光滑的, 根据定理 8.1.1 有

$$H'(x; d) = \lim_{\substack{V \in \partial H(x + td) \\ t \to 0^+}} V d = \lim_{\substack{V \in \partial H(x + t_k d) \\ t_k \to 0^+}} V_k d. \tag{8.1.4}$$

由 V_k 的有界性及极值映射 $\partial H(x)$ 的上半连续性, 则存在 $V \in \partial H(x)$, 使得 $\{V_k\}$ 的一个子列收敛于 V, 由式 (8.1.4) 得

$$H'(x; d) = V d.$$

引理得证.

定理 8.1.2 设 $H(x)$ 为 $\mathbf{R}^n$ 到 $\mathbf{R}^m$ 上的局部 Lipschitz 函数, D_H 为 $H(x)$ 的可微点集, 则下述结论等价:

(1) $H(x)$ 在 x 点半光滑;

(2) 对所有单位向量 d, 式 (8.1.1) 的极限一致收敛;

(3) 对所有单位向量 d, 式 (8.1.2) 的极限一致收敛;

(4) $Vd - H'(x;d) = o(\| d \|), \forall V \in \partial H(x+d)$; (8.1.5)

(5) $\lim\limits_{\substack{x+d\in D_H \\ d\to 0}} \dfrac{H'(x+d;d) - H'(x;d)}{\| d \|} = 0.$ (8.1.6)

证明 (1)⇒(2) 假设结论 (2) 不成立, 则存在常数 $\varepsilon > 0$, 点列 $\{d_k\}, \{\bar{d}_k\}$, 满足

$$\| d_k \| = \| \bar{d}_k \| = 1, \quad \| \bar{d}_k - d_k \| \to 0, \quad t_k \to 0^+ \text{ 和 } V_k \in \partial H(x + t_k\bar{d}_k),$$

使得对任意的 k, 有

$$\| V_k\bar{d}_k - H'(x;d_k) \| \geqslant 2\varepsilon. \tag{8.1.7}$$

选取 $\{d_k\}$ 和 $\{\bar{d}_k\}$ 的收敛子列, 不妨设为 $\{d_k\}$ 和 $\{\bar{d}_k\}$ 本身, 记

$$d_k \to d, \quad \bar{d}_k \to d,$$

注意到局部 Lipschitz 函数方向导数关于方向是连续的, 由式 (8.1.7) 知, 当 k 充分大时有

$$\| V_k\bar{d}_k - H'(x;d) \| \geqslant \varepsilon,$$

这与 $H(x)$ 的半光滑性矛盾.

(2)⇒(3)⇒(4) 考虑引理 8.1.1, 结论显然成立.

(4)⇒(1) 假设 $H(x)$ 在 x 点不是半光滑的, 由式 (8.1.4) 则必存在

$$d \in \mathbf{R}^n, \quad d_k \to d, \quad \varepsilon > 0, \quad t_k \to 0^+, \quad V_k \in \partial H(x + t_k d_k),$$

使得对任意 k 下式成立:

$$\| V_k d_k - H'(x;d) \| \geqslant 2\varepsilon. \tag{8.1.8}$$

再根据 Lipschitz 函数方向导数关于方向的连续性及式 (8.1.8), 对充分大的 k, 有

$$\| V_k d_k - H'(x;d_k) \| \geqslant \varepsilon,$$

这与式 (8.1.5) 矛盾.

(4)⇒(5) 在式 (8.1.5) 中选取

$$V = \nabla H(x+d) \in \partial H(x+d),$$

再考虑到

$$H'(x+d;d) = \nabla H(x+d)^{\mathrm{T}} d,$$

将式 (8.1.5) 代入式 (8.1.6) 左边, 即得结论 (5).

(5)⇒(4) 给定 $\varepsilon > 0$, 从式 (8.1.6) 知, 必存在 $\delta > 0$, 使得对任何满足

$$\|\bar{d}\| \leqslant \delta \quad 和 \quad x + \bar{d} \in D_H$$

的向量 $\bar{d}$, 有

$$\left\|H'(x+\bar{d};\bar{d}) - H'(x;\bar{d})\right\| \leqslant \varepsilon \left\|\bar{d}\right\|. \tag{8.1.9}$$

考虑 $d \in \mathbf{R}^n$ 满足

$$\|d\| \leqslant \frac{1}{2}\delta,$$

以及

$$V \in \partial H(x+d),$$

下面证明

$$\|Vd - H'(x;d)\| \leqslant 5\varepsilon \|d\|. \tag{8.1.10}$$

根据广义 Jacobi 定义及 Jacobi 与方向导数关系得

$$\begin{aligned} Vd \in \ & \mathrm{co}\,\{ \lim_{\substack{d_i \to d \\ x+d_i \in D_H}} JH(x+d_i)d\} \\ = & \mathrm{co}\,\{ \lim_{\substack{d_i \to d \\ x+d_i \in D_H}} H'(x+d_i;d)\}, \end{aligned}$$

再根据 Caratheodory 定理, 必存在 $\bar{d}_0, \cdots, \bar{d}_m$, 使得

$$\left\|\bar{d}_k - d\right\| \leqslant \min\left\{\frac{1}{2}\delta, \|d\|, \frac{\varepsilon}{L}\|d\|\right\}, \quad x + \bar{d}_i \in D_H, \quad k = 0, 1, \cdots, m,$$

$$\left\|Vd - \sum_{k=0}^{m} \lambda_k H'(x+\bar{d}_k;d)\right\| \leqslant \varepsilon, \tag{8.1.11}$$

其中 L 是 $H(x)$ 在 x 处的 Lipschitz 常数,

$$\lambda_k \geqslant 0, \quad k = 1, \cdots, m, \ \ 且 \ \sum_{k=0}^{m} \lambda_k = 1.$$

利用式 (8.1.9) 及函数 $H'(x;\cdot)$ 的 Lipschitz 连续性, 直接推导得

$$\left\|\sum_{k=0}^{m} \lambda_k H'(x+\bar{d}_k;d) - H'(x;d)\right\|$$

$$\leqslant \sum_{k=0}^{m} \lambda_k(\left\|H'(x+\bar{d}_k;\bar{d}_k)-H'(x+\bar{d}_k;d)\right\|+\left\|H'(x+\bar{d}_k;\bar{d}_k)-H'(x;\bar{d}_k)\right\|$$

$$+\left\|H'(x;d)-H'(x;\bar{d}_k)\right\|)$$

$$\leqslant \sum_{k=0}^{m} \lambda_k(L\left\|\bar{d}_k-d\right\|+\varepsilon\left\|\bar{d}_k\right\|+L\left\|\bar{d}_k-d_k\right\|)$$

$$\leqslant \sum_{k=0}^{m} \lambda_k 4\varepsilon\left\|d\right\| = 4\varepsilon\left\|d\right\|, \tag{8.1.12}$$

由式 (8.1.11) 和式 (8.1.12) 得式 (8.1.10), 结论 (4) 成立. 定理得证.

定理 8.1.2 所给出半光滑的 5 个等价条件也可视为其等价定义, 其中条件 (4) 最为常用, 依据 (4) 我们还可给出高价半光滑的定义.

定义 8.1.2 设 $H(x)$ 为 $\mathbf{R}^n$ 到 $\mathbf{R}^m$ 上的局部 Lipschitz 函数且方向可微, 若对于常数 $p\in(0,1]$, 下式成立:

$$Vd-H'(x;d)=O(\|d\|^{1+\rho}),\quad \forall V\in\partial H(x+d),\quad d\in\mathbf{R}^n, \tag{8.1.13}$$

则称 $H(x)$ 是 p 阶半光滑的, 特别当 $p=1$ 时称为是强半光滑的.

显然, 对于 $p\in(0,1]$, p 阶半光滑一定是半光滑的.

命题 8.1.1 设 $H(x)$ 为 $\mathbf{R}^n$ 到 $\mathbf{R}^m$ 上的局部 Lipschitz 函数, 则 $H(x)$ 是半光滑的当且仅当它的每一个分量是半光滑的.

证明 只需考虑充分性. 设

$$H(x)=(h_1(x),\cdots,h_m(x))^{\mathrm{T}},$$

其中每个 $h_i(x)$ 为 $\mathbf{R}^n$ 上的半光滑函数. 根据定理 8.1.2 中结论 (5) 知

$$h_i'(x+d;d)-h_i'(x;d)=o(\|d\|),$$

从而

$$\|H'(x+d;d)-H'(x;d)\|\leqslant\sum_{i=1}^{m}\|h_i'(x+d;d)-h_i'(x;d)\|=o(\|h\|),$$

即 $H(x)$ 是半光滑的. 命题得证.

半光滑与强半光滑下述性质也是常用的, 证明是显然的.

当 $H(x)$ 是半光滑时, 则有

$$H(x+d)-H(x)-H'(x;d)=o(\|d\|); \tag{8.1.14}$$

当 $H(x)$ 是 p 阶半光滑时, 有

$$H(x+d)-H(x)-H'(x;d)=O(\|d\|^{1+\rho}). \tag{8.1.15}$$

8.2　半光滑方程组的牛顿法

本节给出半光滑方程组的牛顿法和不精确牛顿法以及收敛性分析.

8.2.1　牛顿法

考虑下述方程组:

$$F(x)=0, \tag{8.2.1}$$

其中 $F(x)$ 为 $\mathbf{R}^n$ 到 $\mathbf{R}^n$ 上的局部 Lipschitz 函数且是半光滑的. 当然, 我们可以利用各种次微分来建立相应的牛顿法, 例如广义 Jacobi $\partial_{C1}F(x)$、B 微分 $\partial_B F(x)$ 等, 为统一起见, 我们考虑一般形式的次微分, 首先给出关于次微分的一种假设.

假设 8.2.1　设 $F(x)$ 为 $\mathbf{R}^m$ 到 $\mathbf{R}^n$ 上的局部 Lipschitz 函数, 记 $F(x)=(f_1(x),\cdots,f_m(x))^{\mathrm{T}}$, $\partial F(x)$ 为 $F(x)$ 的一种次微分, $x\to\partial F(x)$ 作为 $\mathbf{R}^n$ 到 $\mathbf{R}^{m\times n}$ 子集上的映射是上半连续的, 对每一个 $x\in\mathbf{R}^n$, 集合 $\partial F(x)$ 非空且有

$$\partial F(x)\subset\partial_{C_l}f_1(x)\times\cdots\times\partial_{C_l}f_m(x). \tag{8.2.2}$$

显然, 次微分 $\partial_{C_l}F(x)$, $\partial_B F(x)$ 和 $\partial_{C_l}f_1(x)\times\cdots\times\partial_{C_l}f_m(x)$ 都满足假设 8.2.1.

求解方程组 (8.2.1) 的牛顿法迭代公式如下:

$$x^{k+1}=x^k-V_k^{-1}F(x^k),\quad V_k\in\partial F(x^k), \tag{8.2.3}$$

其中次微分 $\partial F(x)$ 满足假设 8.2.1. 我们可具体选取次微分为广义 Jacobi 和 B 微分等, 选取不同的次微分, 就得到不同的牛顿法. 为证明牛顿法 (8.2.3) 的收敛性, 首先给出一个引理.

引理 8.2.1　设 $F(x)$ 为 $\mathbf{R}^n$ 到 $\mathbf{R}^n$ 上的局部 Lipschitz 函数, 次微分 $\partial F(x)$ 满足假设 8.2.1, 如果所有 $V\in\partial F(x)$ 是非奇异的, 则存在常数 $\beta>0$, 使得

$$\|V^{-1}\|\leqslant\beta,\quad\forall V\in\partial F(x), \tag{8.2.4}$$

进一步存在 x 的邻域 $B(x,\delta)$, 使得

$$\|V^{-1}\|\leqslant\frac{10}{9}\beta,\quad\forall V\in\partial F(x),\quad y\in B(x,\delta). \tag{8.2.5}$$

证明　注意到, 集值映射 $x\to\partial F(x)$ 的上半连续性和 $\partial F(x)$ 的有界性 (由于式 (8.2.2) 右端有界), 易见式 (8.2.4) 和式 (8.2.5) 成立. 引理得证.

定理 8.2.1　设 x^* 是方程组 (8.2.1) 的解, 如果次微分 $\partial F(x)$ 满足假设 8.2.1, $F(x)$ 在 x^* 处是半光滑的以及所有 $V\in\partial F(x^*)$ 都是非奇异的, 则迭代公式 (8.2.3) 产生的点列 $\{x_k\}$ 在点 x^* 附近超线性收敛到 x^*.

证明 根据引理 8.2.1, $\partial F(x^*)$ 中的所有元素是非奇异的, 因此迭代公式 (8.2.3) 在 $k=0$ 是适定的. 记 V_k^i 为 V_k 的第 i 个分量, 由于 $f_i(x), i=1,\cdots,n$ 在 x^* 处是半光滑的,

$$V_k^i \in \partial_{Cl} f_i(x^k),$$

根据半光滑性质有

$$V_k^i(x^k - x^*) - f_i'(x^*; x^k - x^*) = o(\|x^k - x^*\|), \quad i=1,\cdots,n, \tag{8.2.6}$$

进而有

$$V_k(x^k - x^*) - F'(x^*; x^k - x^*) = o(\|x^k - x^*\|). \tag{8.2.7}$$

再由引理 8.2.1, $F(x^*)=0$ 和式 (8.2.7), 推导得

$$\begin{aligned}\|x^{k+1} - x^*\| &= \left\|x^k - x^* - V_k^{-1}F(x^k)\right\| \\ &\leqslant \left\|V_k^{-1}(F(x^k) - F(x^*) - F'(x^*; x^k - x^*))\right\| \\ &\quad + \left\|V_k^{-1}(V_k(x^k - x^*) - F'(x^*; x^k - x^*))\right\| \\ &= o(\left\|x^k - x^*\right\|),\end{aligned}$$

这就证明了在点 x^* 的邻域内超线性收敛到 x^*. 定理得证.

8.2.2 不精确牛顿法

求解非光滑方程组 (8.1.1) 的不精确牛顿法如下:

$$x^{k+1} = x^k - U_k^{-1}F(x^k), \quad U_k \in \mathbf{R}^{n\times n}. \tag{8.2.8}$$

定理 8.2.2 设 x^* 为方程组 (8.1.1) 的解, 次微分 $\partial F(x)$ 满足假设 8.2.1, $F(x)$ 在 x^* 处是半光滑的, 且所有 $V \in \partial F(x^*)$ 是非奇异的, 则存在 $\varepsilon > 0, \Delta > 0$, 如果 $\|x^0 - x^*\| \leqslant \varepsilon$ 且存在 $V^k \in \partial F(x^k)$, 使得

$$\|V_k - U_k\| \leqslant \Delta, \tag{8.2.9}$$

则迭代法 (8.2.8) 是适合的且产生的点列 $\{x^k\}$ 在点 x^* 附近线性收敛到 x^*.

证明 根据引理 8.2.1, 存在 $\beta > 0$ 和 x^* 的邻域 $B(x^*, \delta)$, 使得

$$\|V^{-1}\| \leqslant \beta, \quad \forall V \in \partial F(x^*), \tag{8.2.10}$$

$$\|V^{-1}\| \leqslant \frac{10}{9}\beta, \quad \forall V \in \partial F(y), \quad y \in B(x^*, \delta). \tag{8.2.11}$$

选取 $\Delta > 0$ 满足

$$6\beta\Delta \leqslant 1. \tag{8.2.12}$$

记 $V \in \partial F(x)$ 的第 i 个分量为 V^i, 由 $f_i(x)$ 在 x^* 点的半光滑性以及 $V^i \in \partial_{Cl} f_i(x)$, 则有

$$V^i(x - x^*) - f_i'(x^*; x - x^*) = o\|x - x^*\|, \quad i = 1, \cdots, n,$$

进而得

$$V(x - x^*) - F'(x^*; x - x^*) = o(\|x - x^*\|), \quad \forall V \in \partial F(x). \tag{8.2.13}$$

由式 (8.1.14) 和式 (8.2.13), 选取 ε 充分小, 使得

$$\{x \in \mathbf{R}^n | \|x - x^*\| \leqslant \varepsilon\} \subset B(x^*, \delta),$$

$$\left\|F(x^k) - F(x^*) - V(x^k - x^*)\right\| \leqslant \Delta \|x^k - x^*\|,$$

$$\forall V \in \partial F(x^k), \quad \|x^k - x^*\| \leqslant \varepsilon. \tag{8.2.14}$$

根据式 (8.2.10) 式 (8.2.11) 及 $\|x - x^*\| \leqslant \varepsilon$ 知, $V \in \partial F(x)$ 是非奇异的, 进而有

$$\|V^{-1}\| \leqslant \frac{10}{9}\beta, \quad \forall V \in \partial F(x), \quad \|x - x^*\| \leqslant \varepsilon. \tag{8.2.15}$$

利用矩阵分析中的熟知结果: 设 $A, B \in \mathbf{R}^{n \times n}, B$ 非奇异,

$$\|B^{-1}(B - A)\| < 1,$$

则 A 是非奇异的且有

$$\|A^{-1}\| \leqslant \frac{\|B^{-1}\|}{1 - \|B^{-1}(B - A)\|}. \tag{8.2.16}$$

从现在起, 假设对任意 k 有 $\|x^k - x^*\| \leqslant \varepsilon$, 在式 (8.2.16) 中, 令

$$A = U_k, \quad B = V_k,$$

直接推导得

$$\begin{aligned}\|U_k^{-1}\| &\leqslant \frac{\|V_k^{-1}\|}{1 - \|V_k^{-1}(V_k - U_k)\|} \\ &\leqslant \frac{\frac{10}{9}\beta}{1 - \frac{10}{9}\beta\Delta} \leqslant \frac{\frac{10}{9}\beta}{1 - \frac{5}{27}} \leqslant \frac{3}{2}\beta.\end{aligned} \tag{8.2.17}$$

进一步得

$$\begin{aligned}\|x^{k+1} - x^*\| &= \|x^k - U_k^{-1}F(x^k) - x^*\| \\ &\leqslant \|U_k^{-1}\|\|F(x^k) - F(x^*) - U_k(x^k - x^*)\| \\ &\leqslant \|U_k^{-1}\|\|F(x^k) - F(x^*) - U_k(x^k - x^*)\| \\ &\quad + \|V_k - U_k\|\|x^k - x^*\|.\end{aligned} \tag{8.2.18}$$

将式 (8.2.9), 式 (8.2.12), 式 (8.2.14) 和式 (8.2.17) 代入式 (8.2.18) 得

$$\begin{aligned}\|x^{k+1}-x^*\| &\leqslant \frac{3}{2}\beta(\Delta\|x^k-x^*\|+\Delta\|x^k-x^*\|) \\ &= 3\beta\Delta\|x^k-x^*\| \leqslant \frac{1}{2}\|x^k-x^*\|.\end{aligned} \tag{8.2.19}$$

根据数学归纳法, 由 $\|x^0-x^*\|\leqslant\varepsilon$ 和式 (8.2.19) 则对任意 k, 有

$$\|x^k-x^*\|\leqslant\varepsilon,$$

于是

$$\|x^{k+1}-x^*\|\leqslant\frac{1}{2}\|x^k-x^*\|,$$

这就证明了 $\{x^k\}$ 线性收敛到 x^*. 定理得证.

定理 8.2.3 设 $F(x)$ 在 x^* 处是半光滑的且非奇异, 次微分 $\partial F(x)$ 满足假设 8.2.1, $\{U_k\}$ 为 $\mathbf{R}^{n\times n}$ 中非奇异阵, $\{x^k\}$ 是迭代式 (8.2.8) 产生的点列且 $x^k\neq x^*$, $\lim\limits_{k\to\infty}x^k=x^*$, 则 $\{x^k\}$ 超线性收敛到点 x^* 且 $F(x^*)=0$ 的充要条件是存在 $V_k\in\partial F(x^k)$, 使得

$$\lim_{k\to\infty}\frac{\|(V_k-U_k)s^k\|}{\|s^k\|}=0, \tag{8.2.20}$$

其中 $s^k=x^{k+1}-x^k$.

证明 令 $e^k=x^k-x^*$, 显然点列 $\{e^k\}$ 和 $\{s^k\}$ 都收敛到 0. 根据迭代公式 (8.2.8), 直接推导得

$$\begin{aligned}F(x^*) &= \left(F(x^k)+V_ks^k\right)-\left(F(x^k)-F(x^*)-V_ke^k\right)-V_ke^{k+1} \\ &= F\left(x^k\right)+U_ks^k-(U_k-V_k)\,s^k-\left(F\left(x^k\right)-F\left(x^*\right)-V_ke^k\right)-V_ke^{k+1} \\ &= (U_k-V_k)\,s^k-\left(F\left(x^k\right)-F\left(x^*\right)-V_ke^k\right)-V_ke^{k+1}.\end{aligned} \tag{8.2.21}$$

根据半光滑性质得

$$F\left(x^k\right)-F\left(x^*\right)-V_ke^k=o\left(\|e^k\|\right), \tag{8.2.22}$$

由式 (8.2.20) 得

$$(U_k-V_k)\,s^k=o\left(\|s^k\|\right), \tag{8.2.23}$$

再由式 (8.2.21)~(8.2.23) 及 $e^k\to0$ 和 $\{V_k\}$ 的有界性得

$$F\left(x^*\right)=0,$$

于是

$$(U_k-V_k)\,s^k-\left(F\left(x^k\right)-F\left(x^*\right)-V_ke^k\right)-V_ke^{k+1}=0. \tag{8.2.24}$$

根据引理 8.2.1, $\{V_k^{-1}\}$ 是有界的, 再由式 (8.2.22), 式 (8.2.23) 和式 (8.2.24) 得

$$\begin{aligned}||e^{k+1}|| &\leqslant o(||s^k||) + o(||e^k||)\\ &\leqslant o(||e^k||) + o(||e^{k+1}||),\end{aligned}$$

这说明

$$\lim_{k\to\infty}\frac{||e^{k+1}||}{||e^k||} = 0,$$

即 $\{x_k\}$ 超线性收敛到 x^*. 另一方面, 设 $F(x^*) = 0$, $\{x^k\}$ 超线性收敛到 x^*, 由上面推导的逆过程就可得式 (8.2.20). 定理得证.

8.3 复合函数的牛顿法

8.3.1 牛顿法

考虑复合函数的非光滑方程组:

$$\begin{aligned}&g_1(f_1(x),\cdots,f_m(x)) = 0,\\ &\cdots\cdots\cdots\cdots\\ &g_n(f_1(x),\cdots,f_m(x)) = 0,\end{aligned} \tag{8.3.1}$$

其中 $f_j(x), j = 1,\cdots,m$ 是 $\mathbf{R}^n$ 上的局部 Lipschitz 函数, $g_i(x)$, $i = 1,\cdots,n$ 是 $\mathbf{R}^m$ 上的连续可微函数. 令

$$G(y) = (g_1(y),\cdots,g_n(y))^{\mathrm{T}}, \tag{8.3.2}$$

$$F(x) = (f_1(x),\cdots,f_m(x))^{\mathrm{T}}, \tag{8.3.3}$$

此时方程组 (8.3.1) 可以写成

$$G(F(x)) = 0. \tag{8.3.4}$$

我们知道, 在 8.2 节中给出的牛顿法可以直接应用到方程组 (8.3.1), 然而在某些应用中, 关于函数 $G\circ F(x)$ 次微分的计算是困难的, 而函数 $F(x)$ 次微分的计算却容易, 本节的方法只需函数 $F(x)$ 的次微分, 不需要函数 $G\circ F(x)$ 的次微分.

下面给出解方程组 (8.3.1) 的牛顿法:

$$x^{k+1} = x^k - W_k^{-1}G(F(x^k)), \quad W_k \in JG(F)|_{F=F(x^k)}\partial F(x^k). \tag{8.3.5}$$

一般来说, 尽管

$$\partial F(x) \subset \partial_{Cl}f_1(x)\times\cdots\times\partial_{Cl}f_m(x),$$

但 $JG(F)|_{F=F(x^k)}\partial_F(x^k)$ 却不属于集合

$$\partial_{Cl}g_1(F(x)) \times \cdots \times \partial_{Cl}g_m(F(x)),$$

所以它不是函数 $G \circ F(x)$ 满足假设 8.2.1 的次微分.

引理 8.3.1 令函数 $F(x)$ 和 $G(y)$ 是由 (8.3.2) 和 (8.3.3) 给出的形式, 而 $\partial F(x)$ 是 $F(x)$ 满足假设 8.2.1 的次微分, 如果所有 $W \in JG(F)|_{F=F(x)}\partial F(x)$ 是非奇异的, 那么存在 $\beta > 0$, 使得

$$\|W^{-1}\| \leqslant \beta, \quad \forall W \in JG(F)|_{F=F(x)}\partial F(x), \tag{8.3.6}$$

进一步, 存在 x 的邻域 $B(x,\delta)$, 使得

$$\|W^{-1}\| \leqslant \frac{10}{9}\beta, \quad \forall W \in JG(F)|_{F=F(y)}\partial F(y), \quad y \in B(x,\delta). \tag{8.3.7}$$

证明 注意到函数 $G(y)$ 是连续可微的, 引理的证明可以完全类似于引理 8.2.1.

定理 8.3.1 假设 x^* 是方程组 (8.3.1) 的解, $\partial F(x)$ 是 $F(x)$ 满足假设 8.2.1 的次微分, $F(x)$ 在 x^* 处是半光滑的, 并且所有的 $W \in JG(F)|_{F=F(x^*)}\partial_F(x^*)$ 是非奇异的, 那么式 (8.3.5) 给出的迭代公式是适定的, 其中产生的点列 $\{x^k\}$ 在 x^* 的一个邻域内超线性收敛到 x^*.

证明 根据引理 8.3.1, 迭代公式 (8.3.5) 在 $k=0$ 时是适定的. 记 W_{ki} 为 W_k 的第 i 行, 则 W_{ki} 具有下面的形式:

$$W_{ki} = \sum_{j=1}^{m} \left.\frac{\partial g_i(F)}{\partial f_j}\right|_{F=F(x^k)} V_{kj}, \quad i=1,\cdots,n, \tag{8.3.8}$$

其中

$$V_{k1} \times \cdots \times V_{km} \in \partial F(x^k).$$

根据次微分的定义,

$$V_{kj} \in \partial_{Cl} f_j(x^k), \quad j=1,\cdots,m,$$

再根据式 (8.1.6), 链式法则和 $\dfrac{\partial g_i(F)}{\partial f_j}$ 的局部有界性及连续性有

$$\begin{aligned}
&W_{ki}(x^k - x^*) - (g_i \circ F)'(x^*; x^k - x^*) \\
&= \sum_{j=1}^{m} \left.\frac{\partial g_i(F)}{\partial f_j}\right|_{F=F(x^k)} V_{kj}(x^k - x^*) - \sum_{j=1}^{m} \left.\frac{\partial g_i(F)}{\partial f_j}\right|_{F=F(x^*)} f_j'(x^*; x^k - x^*) \\
&= \sum_{j=1}^{m} \left.\frac{\partial g_i(F)}{\partial f_j}\right|_{F=F(x^k)} (V_{kj}(x^k - x^*) - f_j'(x^*; x^k - x^*))
\end{aligned}$$

$$
\begin{aligned}
&+\sum_{j=1}^{m}\left(\left.\frac{\partial g_i(F)}{\partial f_j}\right|_{F=F(x^k)}-\left.\frac{\partial g_i(F)}{\partial f_j}\right|_{F=F(x^*)}\right)f_j'(x^*;x^k-x^*)\\
&=o(\|x^k-x^*\|),\quad i=1,\cdots,n.
\end{aligned}\tag{8.3.9}
$$

式 (8.3.9) 表明

$$
W_k(x^k-x^*)-(G\circ F)'(x^*;x^k-x^*)=o(\|x^k-x^*\|).\tag{8.3.10}
$$

注意到函数 $G\circ F(x)$ 在 x^* 处是半光滑的, 由式 (8.2.7), 式 (8.3.7) 和式 (8.3.10), 下式成立:

$$
\begin{aligned}
\|x^{k+1}-x^*\|&=\|x^k-x^*-W_k^{-1}G(F(x^k))\|\\
&\leqslant\left\|W_k^{-1}\left(G(F(x^k))-G(F(x^*))-(G\circ F)'(x^*;x^k-x^*)\right)\right\|\\
&\quad+\left\|W_k^{-1}\left(W_k(x^k-x^*)-(G\circ F)'(x^*;x^k-x^*)\right)\right\|\\
&=o(\|x^k-x^*\|),
\end{aligned}
$$

这就证明了 $\{x^k\}$ 超线性收敛到 x^*. 定理得证.

8.3.2　不精确牛顿法

方程组 (8.3.1) 的非精确牛顿法如下:

$$
x^{k+1}=x^k-W_k^{-1}G(F(x^k)),\quad W_k\in JG(F)|_{F=F(x^k)}U_k,\quad U_k\in\mathbf{R}^{n\times n}.\tag{8.3.11}
$$

定理 8.3.2　假设 x^* 处是方程组 (8.3.1) 的解, $\partial F(x)$ 是 $F(x)$ 满足假设 8.2.1 的次微分, $F(x)$ 在 x^* 是半光滑的, 且所有 $W\in JG(F)|_{F=F(x^*)}\partial F(x^*)$ 是非奇异的, 存在 $\varepsilon>0,\Delta>0$, 使得如果 $\|x^0-x^*\|\leqslant\varepsilon$, 并且存在 $V_k\in\partial F(x^k)$, 满足

$$
\|V_k-U_k\|\leqslant\Delta,\tag{8.3.12}
$$

那么式 (8.3.11) 给出的迭代法是适定的且产生的点列 $\{x^k\}$ 线性收敛到 x^*.

证明　由引理 8.3.1 和 $JG(F)|_{F=F(x)}$ 的连续性, 则存在 $\beta>0,\gamma>1$ 和 x^* 的邻域 $B(x^*,\delta)$, 使得

$$
\|W^{-1}\|\leqslant\beta,\quad\forall W\in JG(F)|_{F=F(x^*)}\partial F(x^*),
$$

$$
\|W^{-1}\|\leqslant\frac{10}{9}\beta,\quad\forall W\in JG(F)|_{F=F(y)}V,\quad V\in\partial F(y),\quad y\in B(x^*,\delta),\tag{8.3.13}
$$

$$
\|JG(F)|_{F=F(y)}\|\leqslant\gamma,\quad\forall y\in B(x^*,\delta).\tag{8.3.14}
$$

选择 $\Delta>0$ 满足

$$6\beta\gamma\Delta\leqslant 1. \tag{8.3.15}$$

记 $(JG(F)|_{F=F(x)}V)_i$ 为 $W\in JG(F)|_{F=F(x)}V$ 的第 i 行, 则有

$$(JG(F)|_{F=F(x)}V)_i=\sum_{j=1}^{m}\left.\frac{\partial g_i(F)}{\partial f_j}\right|_{F=F(x)}V_j,$$

其中 $V_1\times\cdots\times V_m\in\partial F(x)$. 注意到在 x^* 处, 每个函数 $g_i(F(x))$ 是半光滑的且 $V_j\in\partial_{Cl}f_j(x)$. 由式 (8.1.4), 式 (8.3.14), V_j 的局部有界性和 $g_i(x)$ 的连续可微性, 对任意 $V\in\partial F(x)$, 下式成立:

$$\begin{aligned}
&\|g_i(F(x))-g_i(F(x^*))-(JG(F)|_{F=F(x)}V)_i(x-x^*)\|\\
=&\left\|g_i(F(x))-g_i(F(x^*))-\sum_{j=1}^{m}\left.\frac{\partial g_i(F)}{\partial f_j}\right|_{F=F(x)}V_j(x-x^*)\right\|\\
\leqslant&\|g_i(F(x))-g_i(F(x^*))-(g_i\circ F)'(x^*;x-x^*)\|\\
&+\left\|(g_i\circ F)'(x^*;x-x^*)-\sum_{j=1}^{m}\left.\frac{\partial g_i(F)}{\partial f_j}\right|_{F=F(x)}V_j(x-x^*)\right\|\\
=&o(\|x-x^*\|)+\left\|\sum_{j=1}^{m}\left.\frac{\partial g_i(F)}{\partial f_j}\right|_{F=F(x^*)}f_j'(x^*;x-x^*)\right.\\
&\left.-\sum_{j=1}^{m}\left.\frac{\partial g_i(F)}{\partial f_j}\right|_{F=F(x)}V_j(x-x^*)\right\|\\
\leqslant&o(\|x-x^*\|)+\left\|\sum_{j=1}^{m}\left.\frac{\partial g_i(F)}{\partial f_j}\right|_{F=F(x^*)}[f_j'(x^*;x-x^*)-V_j(x-x^*)]\right\|\\
&+\sum_{j=1}^{m}\left\|\left[\left.\frac{\partial g_i(F)}{\partial f_j}\right|_{F=F(x)}-\left.\frac{\partial g_i(F)}{\partial f_j}\right|_{F=F(x)}\right]V_j(x-x^*)\right\|\\
=&o(\|x-x^*\|),\quad i=1,\cdots,n.
\end{aligned}\tag{8.3.16}$$

因此, 对任意 $V\in\partial F(x)$, 下式成立:

$$\|G(F(x))-G(F(x^*))-JG(F)|_{F=F(x)}V(x-x^*)\|=o(\|x-x^*\|). \tag{8.3.17}$$

根据式 (8.3.17), 选取足够小的 $\varepsilon>0$, 使得

$$\{x\in\mathbf{R}^n|\,\|x-x^*\|\leqslant\varepsilon\}\subset B(x^*,\delta),$$

$$\left\|G(F(x)) - G(F(x^*)) - JG(F)|_{F=F(x)} V(x - x^*)\right\| \leqslant \Delta(\|x - x^*\|),$$
$$\forall V \in \partial F(x), \quad \|x - x^*\| \leqslant \varepsilon. \tag{8.3.18}$$

由式 (8.3.13) 和式 (8.3.14), $\|x - x^*\| \leqslant \varepsilon$ 意味着对任意 $V \in \partial F(x)$, $JG(F)|_{F=F(x)} V$ 是非奇异的且满足

$$\|(JG(F)|_{F=F(x)} V)^{-1}\| \leqslant \frac{10}{9}\beta, \tag{8.3.19}$$

$$\left\|JG(F)|_{F=F(x)}\right\| \leqslant \gamma. \tag{8.3.20}$$

下面假设 $\|x - x^*\| \leqslant \varepsilon$. 在式 (8.2.16) 中, 令

$$A = W_k = JG(F)|_{F=F(x^k)} U_k, \quad B = JG(F)|_{F=F(x^k)} V_k,$$

于是有

$$\begin{aligned}\left\|W_k^{-1}\right\| &\leqslant \frac{\left\|(JG(F)|_{F=F(x^k)} V_k)^{-1}\right\|}{1 - \left\|(JG(F)|_{F=F(x^k)} V_k)^{-1}(JG(F)|_{F=F(x^k)}(V_k - U_k))\right\|} \\ &\leqslant \frac{\frac{10}{9}\beta}{1 - \frac{10}{9}\beta\gamma\Delta} \leqslant \frac{\frac{10}{9}\beta}{1 - \frac{5}{27}} \leqslant \frac{3}{2}\beta.\end{aligned} \tag{8.3.21}$$

直接推导得

$$\begin{aligned}\left\|x^{k+1} - x^*\right\| &= \left\|x^k - W_k^{-1} G(F(x^k)) - x^*\right\| \\ &\leqslant \left\|W_k^{-1}\right\| \left\|G(F(x^k)) - G(F(x^*)) - W_k(x^k - x^*)\right\| \\ &\leqslant \left\|W_k^{-1}\right\| \left[\left\|G(F(x^k)) - G(F(x^*)) - JG(F)|_{F=F(x^k)} V_k(x^k - x^*)\right\|\right. \\ &\quad \left. + \left\|JG(F)|_{F=F(x^k)}(V_k - U_k)\right\| \left\|(x^k - x^*)\right\|\right].\end{aligned} \tag{8.3.22}$$

将式 (8.3.15), 式 (8.3.18), 式 (8.3.20) 和式 (8.3.21) 代入式 (8.3.22), 得

$$\begin{aligned}\left\|x^{k+1} - x^*\right\| &\leqslant \frac{3}{2}\beta(\Delta\|x^k - x^*\| + \gamma\Delta\|x^k - x^*\|) \\ &= \frac{3}{2}(\beta\Delta + \beta\gamma\Delta)\left\|x^k - x^*\right\| \\ &\leqslant \frac{1}{4}\left(\frac{1}{\gamma} + 1\right)\left\|x^k - x^*\right\| \\ &\leqslant \frac{1}{2}\left\|x^k - x^*\right\|.\end{aligned} \tag{8.3.23}$$

由数学归纳法, $\|x^0 - x^*\| \leqslant \varepsilon$ 和式 (8.3.23) 得到

$$\|x^k - x^*\| \leqslant \varepsilon$$

对所有 k 成立, 因此在 $\|x^0 - x^*\| \leqslant \varepsilon$ 的假设下, 关系式

$$\|x^{k+1} - x^*\| \leqslant \frac{1}{2}\|x^k - x^*\|$$

对所有 k 成立, 这说明迭代式 (8.3.11) 是适定的, 且序列 $\{x_k\}$ 线性收敛到 x^*. 定理得证.

例 8.3.1 考虑非光滑方程

$$\begin{aligned} f_1(x) + f_2(x) &= 0, \\ f_3(x)f_4(x) &= 0, \end{aligned} \tag{8.3.24}$$

其中 $f_j(x), j = 1, \cdots, 4$ 是 $\mathbf{R}^2$ 上的局部 Lipschitz 函数, 且在方程组 (8.3.24) 的解 x^* 处是半光滑的, 利用牛顿法 (8.3.5) 解方程组 (8.3.24). 令

$$\begin{aligned} F(x) &= (f_1(x), f_2(x), f_3(x), f_4(x))^{\mathrm{T}}, \\ g_1(f_1, f_2, f_3, f_4) &= f_1 + f_2, \\ g_2(f_1, f_2, f_3, f_4) &= f_3 f_4, \\ G(y) &= (g_1(y), g_2(y))^{\mathrm{T}}, \end{aligned}$$

可得

$$JG(F) = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & f_4 & f_3 \end{pmatrix}.$$

给定

$$V_k = V_{k1} \times V_{k2} \times V_{k3} \times V_{k4} \in \partial F(x^k),$$

有

$$\begin{aligned} W_k &= JG(F)|_{F=F(x^k)} V_k \\ &= \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & f_4(x^k) & f_3(x^k) \end{pmatrix} \begin{pmatrix} V_{k1} \\ V_{k2} \\ V_{k3} \\ V_{k4} \end{pmatrix} \\ &= \begin{pmatrix} V_{k1} + V_{k2} \\ f_4(x^k)V_{k3} + f_3(x^k)V_{k4} \end{pmatrix}, \end{aligned}$$

因此, 解方程组 (8.3.24) 牛顿法的迭代公式是

$$x^{k+1} = x^k + \begin{pmatrix} V_{k1} + V_{k2} \\ f_4(x^k)V_{k3} + f_3(x^k)V_{k4} \end{pmatrix}^{-1} \begin{pmatrix} f_1(x^k) + f_2(x^k) \\ f_3(x^k)f_4(x^k) \end{pmatrix}.$$

8.4 拟可微方程组的牛顿法

前两节讨论的非光滑方程组牛顿法是针对局部 Lipschitz 函数, 利用 Clarke 广义 Jacobi 型微分建立的, 本节讨论另一类非光滑函数、拟可微函数的方程组、基于拟微分建立相应的牛顿法.

8.4.1 拟半光滑

为建立收敛性理论, 首先引入拟半光滑概念.

设 $F(x)$ 为 $\mathbf{R}^n$ 到 $\mathbf{R}^m$ 上的函数, $f_i(x)$ 为 $F(x)$ 的第 i 个分量, $f_i(x)$ 是拟可微的, 其拟微分为 $[\underline{\partial} f_i(x), \bar{\partial} f_i(x)]$, 则称 $F(x)$ 是拟可微的, 其拟微分定义为

$$[\underline{\partial} F(x), \bar{\partial} F(x)] = [\underline{\partial} f_1(x), \bar{\partial} f_1(x)] \times \cdots \times [\underline{\partial} f_m(x), \bar{\partial} f_m(x)], \tag{8.4.1}$$

显然 $\underline{\partial} F(x)$ 和 $\bar{\partial} F(x)$ 为 $\mathbf{R}^{n\times m}$ 中的凸紧集.

在 $m \times n$ 维矩阵空间 $\mathbf{R}^{m\times n}$ 中的凸紧集 W 的支撑函数 $\delta^*(W; \cdot)$ 为

$$\delta^*(x|W) = \max_{\xi\in W} \xi x,\ x \in \mathbf{R}^n.$$

设 $F(x)$ 为 $\mathbf{R}^n$ 到 $\mathbf{R}^m$ 上的拟可微函数, $[\underline{\partial} F(x), \bar{\partial} F(x)]$ 为拟微分, 则有

$$F'(x; d) = \delta^*(d|\underline{\partial} F(x)) - \delta^*(d| - \bar{\partial} F(x)), \quad d \in \mathbf{R}^n.$$

$F(x)$ 为 $\mathbf{R}^n$ 到 $\mathbf{R}^m$ 上的函数, 如果

$$F(x+d) - F(x) - F'(x; d) = o(\|d\|), \quad d \in \mathbf{R}^n, \tag{8.4.2}$$

则称 $F(x)$ 在 $x \in \mathbf{R}^n$ 处是 B 可微的 (一致方向可微).

定义 8.4.1　设 $S(x)$ 为 $\mathbf{R}^n$ 到 $\mathbf{R}^{m\times n}$ 中子集上的集值映射, 如果

$$\xi d - \delta^*(d|S(x)) = o(\|d\|), \quad \forall \xi \in S(x+d)$$

成立, 则称 $S(x)$ 在 x 处是拟方向连续的.

容易证明, 集值映射 $\partial_{Cl} F(\cdot)$ 是拟方向连续的当且仅当 $F(x)$ 是半光滑的. 注意到

$$\delta^*(d|S_1(x) + S_2(x)) = \delta^*(d|S_1(x)) + \delta^*(d|S_2(x)),$$

$$\delta^*(d|\lambda S(x)) = \lambda\delta^*(d|S(x)), \quad \lambda \geqslant 0,$$

我们有以下命题.

命题 8.4.1 假设 $\mathbf{R}^n$ 到 $\mathbf{R}^{m\times n}$ 的子集上的集值映射 $S_1(x)$ 和 $S_2(x)$ 是拟方向连续的, $\lambda \geqslant 0$, 那么 $S_1(x)+S_2(x)$ 和 $\lambda S_2(x)$ 也是拟方向连续的.

定义 8.4.2 设 $F(x)$ 为 $\mathbf{R}^n$ 到 $\mathbf{R}^m$ 上的拟可微函数, 拟微分为 $[\underline{\partial}F(x),\bar{\partial}F(x)]$, 如果集值映射 $\underline{\partial}F(\cdot)$ 和 $-\bar{\partial}F(\cdot)$ 在 x 处是拟方向可微的, 即

$$Ud-\delta^*(d|\underline{\partial}F(x))=o(\|d\|),\quad U\in\underline{\partial}F(x+d),$$

$$Vd-\delta^*(d|-\bar{\partial}F(x))=o(\|d\|),\quad V\in\bar{\partial}F(x+d),$$

称 $F(x)$ 在点 x 关于拟微分 $[\underline{\partial}F(x),\bar{\partial}F(x)]$ 是拟半光滑的,

注记 8.4.1 令 $f_i(x)$ 为 $\mathbf{R}^n$ 到 $\mathbf{R}$ 上的拟可微函数, 拟微分为 $[\underline{\partial}f_i(x),\bar{\partial}f_i(x)]$, 设 $F(x)=(f_1(x),\cdots,f_m(x))^{\mathrm{T}}$ 的拟微分如式 (8.4.1), 那么 $F(x)$ 关于拟微分 $[\underline{\partial}F(x),\bar{\partial}F(x)]$ 是拟半光滑的当且仅当所有的 $f_i(x)$ 关于拟微分 $[\underline{\partial}f_i(x),\bar{\partial}f_i(x)]$, $i=1,\cdots,m$ 是拟半光滑的.

根据拟微分的运算法则和定理 8.4.1, 如果 $F_1(x)$ 和 $F_2(x)$ 都是拟半光滑的, 则 $F_1(x)+F_2(x)$ 是拟半光滑的.

定理 8.4.1 假设 $f_i(x)$ 为 $\mathbf{R}^n$ 到 $\mathbf{R}$ 上的半光滑函数且有 $f_i^\circ(x;d)=f_i'(x;d), i=1,2$, 那么函数 $f(x)=f_1(x)-f_2(x)$ 和 $g(x)=f_1(x)+f_2(x)$ 是拟半光滑的.

证明 由

$$f_i^\circ(x;d)=f_i'(x;d),$$

则有

$$f'(x;d)=\delta^*(d|\partial_{Cl}f_1(x))-\delta^*(d|-\partial_{Cl}f_2(x)),$$

故 $f(x)$ 是拟可微的, 其拟微分为

$$[\underline{\partial}f(x),\bar{\partial}f(x)]=[\partial_{Cl}f_1(x),-\partial_{Cl}f_2(x)].$$

因为 $f_1(x)$ 和 $f_2(x)$ 都是半光滑的, 则有

$$Ud-\delta^*(d|\partial_{Cl}f_1(x))=o(\|d\|),\quad U\in\partial_{Cl}f_1(x+d), \tag{8.4.3}$$

$$Vd-\delta^*(d|\partial_{Cl}f_2(x))=o(\|d\|),\quad V\in\partial_{Cl}f_2(x+d), \tag{8.4.4}$$

于是

$$\underline{\partial}f(\cdot)=\partial_{Cl}f_1(\cdot),\quad -\bar{\partial}f(\cdot)=\partial_{Cl}f_2(\cdot)$$

是拟方向可微的, 因此 $f(x)$ 是拟半光滑的. 容易看到

$$\begin{aligned}g'(x;d)=&\delta^*(d|\partial_{Cl}f_1(x))+\delta^*(d|\partial_{Cl}f_2(x))\\=&\delta^*(d|\partial_{Cl}f_1(x)+\partial_{Cl}f_2(x)),\end{aligned}$$

因此 $g(x)$ 是拟可微的, 拟微分为

$$\left[\underline{\partial} g(x), \bar{\partial} g(x)\right] = \left[\partial_{Cl} f_1(x) + \partial_{Cl} f_2(x), \{0\}\right].$$

由于集值映射 $\partial_{Cl} f_1(\cdot)$ 和 $\partial_{Cl} f_2(\cdot)$ 是拟方向连续的, 据定理 8.4.1,

$$\underline{\partial} g(\cdot) = \partial_{Cl} f_1(\cdot) + \partial_{Cl} f_2(\cdot)$$

是拟方向连续的, 另一方面,

$$-\bar{\partial} g(x) = \{0\}$$

是拟方向连续的, 根据定义, $g(x)$ 是拟半光滑的.

8.4.2 牛顿法

考虑下面的方程组:

$$H(x) = 0, \tag{8.4.5}$$

其中 $H(x)$ 为 $\mathbf{R}^n$ 到 $\mathbf{R}^n$ 上的拟可微函数. 求解方程组 (8.4.5) 的牛顿法如下:

$$x^{k+1} = x^k - (U_k + V_k)^{-1} H(x^k), \quad [U_k, V_k] \in [\underline{\partial} H(x^k), \bar{\partial} H(x^k)]. \tag{8.4.6}$$

引理 8.4.1 假设集值映射 $\underline{\partial} H(\cdot)$ 和 $\bar{\partial} H(\cdot)$ 在 x 处是上半连续的, 如果所有的 $U + V \in \underline{\partial} H(\cdot) + \bar{\partial} H(\cdot)$ 是非奇异的, 那么存在 $\beta > 0$, 使得

$$\|(U + V)^{-1}\| \leqslant \beta, \quad \forall [U, V] \in [\underline{\partial} H(x), \bar{\partial} H(x)]. \tag{8.4.7}$$

此外, 存在 x 的邻域 $B(x, \delta)$, 使得

$$\|(U+V)^{-1}\| \leqslant \frac{10}{9}\beta, \quad \forall [U, V] \in [\underline{\partial} H(y), \bar{\partial} H(y)], \quad \forall y \in B(x, \delta). \tag{8.4.8}$$

证明 因为 $\underline{\partial} H(\cdot)$ 和 $\bar{\partial} H(\cdot)$ 是上半连续的, 所以 $\underline{\partial} H(\cdot) + \bar{\partial} H(\cdot)$ 也是上半连续, 注意到 $\underline{\partial} H(x) + \bar{\partial} H(x)$ 是有界的, 因此式 (8.4.7) 和式 (8.4.8) 成立.

定理 8.4.2 假设 x^* 是 (8.4.5) 的解, 所有的 $U + V \in \underline{\partial} H(x^*) + \bar{\partial} H(x^*)$ 是非奇异的, $H(x)$ 在 x^* 处是拟半光滑且 B 可微的, 集值映射 $\underline{\partial} H(\cdot)$ 和 $\bar{\partial} H(\cdot)$ 在 x^* 处是上半连续的, 那么式 (8.4.6) 给出的迭代法是适定的且点列 $\{x^k\}$ 在 x^* 的邻域内超线性收敛到 x^*.

证明 由引理 8.4.1, 当 $k = 0$ 时迭代法 (8.4.6) 在 x^* 的邻域内是适定的. 由于 $H(x)$ 在 x^* 处是拟半光滑的, 根据定义 8.4.2, 有

$$\begin{aligned}
&(U + V)(x - x^*) - H'(x^*; x - x^*) \\
&= U(x - x^*) - \delta^*(x - x^* | \underline{\partial} H(x^*)) + V(x - x^*) + \delta^*(x - x^* | - \bar{\partial} H(x^*)) \\
&= o(\|x - x^*\|), \quad [U, V] \in [\underline{\partial} H(x), \bar{\partial} H(x)].
\end{aligned} \tag{8.4.9}$$

由引理 8.4.1 和式 (8.4.9), $H(x)$ 在 x^* 处的拟半光滑性、B 可微性和 $H(x^*)=0$, 推到得

$$\begin{aligned}\|x^{k+1}-x^*\| &= \|x^k-x^*-(U_k+V_k)^{-1}H(x^k)\| \\ &\leqslant \|(U_k+V_k)^{-1}(H(x^k)-H(x^*)-H'(x^*;x^k-x^*))\| \\ &\quad +\|(U_k+V_k)^{-1}((U_k+V_k)(x^k-x^*)-H'(x^*;x^k-x^*))\| \\ &\quad -o(\|x^k-x^*\|).\end{aligned}$$

这表明点列 $\{x^k\}$ 超线性收敛到 x^*. 定理得证.

8.4.3 不精确牛顿法

求解 (8.4.5) 方程组的非精确的牛顿法如下:

$$x^{k+1}=x^k-W_k^{-1}H(x^k),\quad W_k\in\mathbf{R}^{n\times n}. \tag{8.4.10}$$

定理 8.4.3 假设 x^* 是方程组 (8.4.5) 的解, 定理 8.4.2 的假设条件成立, 存在 $\varepsilon,\Delta>0$, 如果 $\|x^0-x^*\|\leqslant\varepsilon$ 且存在 $U_k\in\underline{\partial}H(x^k)$ 和 $V_k\in\bar{\partial}H(x^k)$, 使得

$$\|U_k+V_k-W_k\|\leqslant\Delta, \tag{8.4.11}$$

那么, (8.4.10) 中的迭代法是适定的且产生的点列 $\{x^k\}$ 在点 x^* 附近线性收敛到 x^*.

证明 由引理 8.4.1, 存在 $\beta>0$ 和 x^* 的邻域 $B(x^*,\delta)$, 使得

$$\|(U+V)^{-1}\|\leqslant\frac{10}{9}\beta,\quad \forall\,[U,V]\in\left[\underline{\partial}H(y),\bar{\partial}H(y)\right],\quad y\in B(x^*,\delta), \tag{8.4.12}$$

选择 $\Delta>0$ 满足

$$6\beta\Delta\leqslant 1. \tag{8.4.13}$$

注意到 $H(x)$ 在 x^* 处是 B 可微且是拟半光滑的, 根据式 (8.4.2) 和式 (8.4.9), 有

$$\begin{aligned}&H(x)-H(x^*)-(U+V)(x-x^*)\\ &=(H(x)-H(x^*)-H'(x^*;x-x^*))\\ &\quad +H'(x^*;x-x^*)-(U+V)(x-x^*)\\ &=o(\|x-x^*\|),\quad [U,V]\in[\underline{\partial}H(x),\bar{\partial}H(x)].\end{aligned} \tag{8.4.14}$$

在式 (8.4.14) 中, 选取足够小的 $\varepsilon>0$, 使得

$$\{x\in\mathbf{R}^n \mid \|x-x^*\|\leqslant\varepsilon\}\subset B(x^*,\delta), \tag{8.4.15}$$

$$\|H(x) - H(x^*) - (U+V)(x-x^*)\| \leqslant \Delta \|x - x^*\|,$$
$$[U,V] \in \left[\underline{\partial} H(x), \bar{\partial} H(x)\right], \quad \|x - x^*\| \leqslant \varepsilon. \tag{8.4.16}$$

由式 (8.4.12) 和式 (8.4.15), 得

$$\|(U+V)^{-1}\| \leqslant \frac{10}{9}\beta, \quad \forall [U,V] \in [\underline{\partial} H(x), \bar{\partial} H(x)], \quad \|x - x^*\| \leqslant \varepsilon. \tag{8.4.17}$$

假设

$$\|x - x^*\| \leqslant \varepsilon,$$

在式 (8.2.16) 中, 令

$$A = W_k, \quad B = U_k + V_k,$$

由式 (8.4.11)、式 (8.4.13) 和式 (8.4.17) 推导得

$$\begin{aligned}
\|W_k^{-1}\| &\leqslant \frac{\|(U_k+V_k)^{-1}\|}{1 - \|(U_k+V_k)^{-1}(U_k+V_k-W_k)\|} \\
&\leqslant \frac{\frac{10}{9}\beta}{1 - \frac{10}{9}\beta\Delta} \leqslant \frac{\frac{10}{9}\beta}{1 - \frac{5}{27}} \leqslant \frac{3}{2}\beta.
\end{aligned} \tag{8.4.18}$$

这意味着 W_k 是非奇异的, 因此迭代公式 (8.4.10) 是适定的. 直接推导得

$$\begin{aligned}
\|x^{k+1} - x^*\| &= \|x^k - W_k^{-1} H(x^k) - x^*\| \\
&= \|W_k^{-1}\| \, \|H(x^k) - H(x^*) - W_k(x^k - x^*)\| \\
&\leqslant \|W_k^{-1}\| (\|H(x^k) - H(x^*) - (U_k+V_k)(x^k - x^*)\| \\
&\quad + \|V_k + U_k - W_k\| \|x^k - x^*\|).
\end{aligned} \tag{8.4.19}$$

将式 (8.4.11), 式 (8.4.13), 式 (8.4.16) 和式 (8.4.18) 代入式 (8.4.19) 得

$$\begin{aligned}
\|x^{k+1} - x^*\| &\leqslant \frac{3}{2}\beta(\Delta\|x^k - x^*\| + \Delta\|x^k - x^*\|) \\
&= 3\beta\Delta\|x^k - x^*\| \\
&\leqslant \frac{1}{2}\|x^k - x^*\|.
\end{aligned} \tag{8.4.20}$$

于是, 由 $\|x^0 - x^*\| \leqslant \varepsilon$ 和式 (8.4.20), 得

$$\|x^k - x^*\| \leqslant \varepsilon$$

对任意 k 成立, 因此在 $\|x^0 - x^*\| \leqslant \varepsilon$ 的假设下, 对任意 k 有

$$\|x^{k+1} - x^*\| \leqslant \frac{1}{2}\|x^k - x^*\|,$$

即序列 $\{x^k\}$ 线性收敛到 x^*. 定理得证.

下面的定理给出了不精确牛顿法 (8.4.10) 超线性收敛的充分必要条件.

定理 8.4.4 假设所有 $U+V\in\underline{\partial}H(x^*)+\bar{\partial}H(x^*)$ 是非奇异的, $H(x)$ 在 x^* 处是拟半光滑且 B 可微的, 集值映射 $\underline{\partial}H(\cdot)$ 和 $\bar{\partial}H(\cdot)$ 在 x^* 处是上半连续的, $\{W_k\}$ 是 $m\times n$ 维矩阵序列, $\{x_k\}$ 是由迭代法 (8.4.10) 产生的点列且对任意 k 有 $x^k\neq x^*$, $\lim\limits_{k\to\infty}x^k=x^*$, 那么 $\{x^k\}$ 超线性收敛到 x^* 且有 $H(x^*)=0$ 当且仅当存在 $U_k\in\underline{\partial}H(x^k)$, $V_k\in\bar{\partial}H(x^k)$, 使得

$$\lim_{k\to\infty}\frac{\|(U_k+V_k-W_k)(x^{k+1}-x^k)\|}{\|x^{k+1}+x^k\|}=0. \tag{8.4.21}$$

证明 充分性. 假设 (8.4.21) 成立, 令

$$e^k=x^k-x^*,\quad s^k=x^{k+1}-x^k,$$

因此点列 $\{e^k\}$ 和 $\{x^k\}$ 收敛到 0. 由式 (8.4.10) 可知

$$H(x^k)+W_ks^k=0.$$

直接推导得

$$\begin{aligned}H(x^*)=&(H(x^k)+(U_k+V_k)s^k)-(U_k+V_k)e^{k+1}\\&-(H(x^k)-H(x^*)-(U_k+V_k)e^k)\\=&H(x^k)+W_ks^k+(U_k+V_k-W_k)s^k-(U_k+V_k)e^{k+1}\\&-(H(x^k)-H(x^*)-(U_k+V_k)e^k)\\=&(U_k+V_k-W_k)s^k-(U_k+V_k)e^{k+1}\\&-(H(x^k)-H(x^*)-(U_k+V_k)e^k).\end{aligned} \tag{8.4.22}$$

联立式 (8.4.14) 和 (8.4.21), 则有

$$H(x^k)-H(x^*)-(U_k+V_k)e^k=o(\|e^k\|), \tag{8.4.23}$$

$$(U_k+V_k-W_k)s^k=o(\|s^k\|). \tag{8.4.24}$$

因此, 由式 (8.4.22)~(8.4.24), $e^k\to0$ 和 $\{U_k+V_k\}$ 的有界性有

$$H(x^*)=0,$$

于是

$$(U_k+V_k-W_k)s^k-(U_k+V_k)e^{k+1}-(H(x^k)-H(x^*)-(U_k+V_k)e^k)=0. \tag{8.4.25}$$

由式 (8.4.23)~(8.4.25) 和 $\{(U_k+V_k)^{-1}\}$ 的有界性, 下式成立:

$$\begin{aligned}\|e^{k+1}\| &\leqslant o(\|s^k\|)+o(\|e^k\|)\\ &\leqslant o(\|e^k\|)+o(\|e^{k+1}\|),\end{aligned}$$

这意味着

$$\lim_{k\to\infty}\frac{\|e^{k+1}\|}{\|e^k\|}=0,$$

即 $\{x^k\}$ 超线性收敛到 x^*.

必要性. 假设 $H(x^*)=0$ 且 $\{x^k\}$ 超线性收敛到 x^*. 由以上讨论的逆过程中就可以直接得到式 (8.4.21). 定理得证.

8.4.4 举例

下面给出两个拟可微方程组的例子, 主要讨论相关函数的拟半光滑性.

例 8.4.1　考虑函数

$$P(x)=\Phi(x)-\Psi(x), \tag{8.4.26}$$

其中 $\Phi(x)$ 和 $\Psi(x)$ 都是为 $\mathbf{R}^n$ 到 $\mathbf{R}^n$ 的凸函数, 也就是说, 它们的所有分量都是凸函数. 记 $\Phi(x)$, $\Psi(x)$, $P(x)$ 的第 i 分量分别为 $\phi_i(x),\psi_i(x),p_i(x)$, 显然 $P(x)$ 是拟可微的且它的拟微分是

$$\begin{aligned}\underline{\partial}P(x)&=\partial\phi_1(x)\times\cdots\times\partial\phi_n(x),\\ \bar{\partial}P(x)&=(-\partial\psi_1(x)\times\cdots\times(-\partial\psi_n(x)),\end{aligned}$$

其中 ∂ 表示凸函数的次微分. 集值映射 $\partial\phi_i(\cdot)$ 和 $\partial\psi_i(\cdot)$ 是上半连续的, 因此 $\underline{\partial}P(\cdot)$ 和 $\bar{\partial}P(\cdot)$ 也是上半连续的. 注意到 $\phi_i^\circ(\cdot;d)=\phi_i'(\cdot;d)$ 且 $\psi_i^\circ(\cdot;d)=\psi_i'(\cdot;d)$, 同时 $\phi_i(x)$ 和 $\psi_i(x)$, $i=1,\cdots,m$ 是半光滑函数, 根据定理 8.4.2, $p_i(x)$ 是拟半光滑的, 因此 $P(x)$ 也是拟半光滑的.

例 8.4.2　考虑下面的函数:

$$Q(x)=\max_{i\in I}F_i(x)-\max_{j\in J}G_j(x), \tag{8.4.27}$$

其中 $F_i(x)$ 和 $G_j(x)$ 为 $\mathbf{R}^n$ 到 $\mathbf{R}^n$ 的连续可微函数, I 和 J 为有界指标集. 不失一般性, 假设函数 $F_i(x)$ 和 $G_j(x)$ 为 $\mathbf{R}^n$ 到 $\mathbf{R}$ 上的函数. 记

$$\begin{aligned}I(x)&=\{i\in I|F_i(x)=\max_{i\in I}F_i(x)\},\\ J(x)&=\{j\in J|G_j(x)=\max_{j\in J}G_j(x)\},\end{aligned}$$

显然 $Q(x)$ 是拟可微的, 它的拟微分为

$$\underline{\partial} Q(x) = \mathrm{co}\{\nabla F_i(x) | i \in I(x)\},$$

$$\bar{\partial} Q(x) = \mathrm{co}\{\nabla G_j(x) | j \in J(x)\}.$$

令

$$Q_1(x) = \max_{i \in I} F_i(x), \quad Q_2(x) = \max_{j \in J} G_j(x),$$

于是

$$Q_i^{\circ}(\cdot; d) = Q_i'(\cdot; d), \quad i = 1, 2.$$

因为 $Q_1(x)$ 和 $Q_2(x)$ 是半光滑的, 根据定理 8.4.1, $Q(x)$ 是拟半光滑的.

8.5 非线性互补问题

本节讨论将一个非线性互补问题转化为非光滑方程组, 这样就可以用上节介绍非光滑方程组的牛顿法来求解.

8.5.1 互补问题与非线性与函数

非线性互补问题就是求解下述问题:

$$f(x) \geqslant 0, \quad h(x) \geqslant 0, \quad f(x)^{\mathrm{T}} h(x) = 0, \tag{8.5.1}$$

其中

$$f(x) = (f_1(x), \cdots, f_n(x))^{\mathrm{T}}, \quad h(x) = (h_1(x), \cdots, h_n(x))^{\mathrm{T}},$$

$$f_i(x), \quad h_i(x), \quad i = 1, \cdots, n$$

均为 $\mathbf{R}^n$ 上的连续可微函数. 非线性互补问题有广泛的应用, 非线性规划的 Kuhn-Tucker 系统就是一个典型的非线性互补问题. 所谓非线性互补函数就是能够将问题 (8.5.1) 转化为一个非光滑方程组的函数.

定义 8.3.1 设 $\phi(a, b)$ 为 $\mathbf{R}^2$ 上的实函数, 如果它满足下述性质:

$$\phi(a, b) = 0 \Leftrightarrow a \geqslant 0, b \geqslant 0, ab = 0,$$

则称 $\phi(a, b)$ 为一个非线性互补函数, 简记为 NCP 函数.

利用非线性互补函数 $\phi(a, b)$, 非线性互补问题 (8.5.1) 等价于下述方程组:

$$\phi(f_i(x), h_i(x)) = 0, \quad i = 1, \cdots, n. \tag{8.5.2}$$

最常见的两个 NCP 函数为极小算子和 Fischer-Burmeister 的 NCP 函数. 不难验证, 下述结论成立:

$$\min\{a,b\}=0 \Leftrightarrow a\geqslant 0, b\geqslant 0, ab=0,$$

$$\sqrt{a^2+b^2}-a-b \Leftrightarrow a\geqslant 0, b\geqslant 0, ab=0.$$

利用这两个非线性互补函数, 非线性互补问题 (8.5.1) 等价于下述两个非光滑方程组

$$\min\{\, f_i(x)\,,\, h_i(x)\,\}=0, \quad i=1,\cdots,n, \tag{8.5.3}$$

$$\sqrt{x_i^2+f_i^2(x)}-x_i-f_i(x), \quad i=1,\cdots,n. \tag{8.5.4}$$

显然, 方程组 (8.5.3) 和方程组 (8.5.4) 是非光滑的, 相应的函数是局部 Lipschitz 和半光滑的, 因此本章前面介绍的牛顿法可用于求解上述两个问题.

8.5.2　非线性互补函数性质

命题 8.5.1　Fischer-Burmeister 非线形互补函数

$$\phi(a,b)=\sqrt{a^2+b^2}-a-b \tag{8.5.5}$$

在原点以外是连续可微的, 在原点处是局部 Lipschitz 和强半光滑的, 其在原点的 B 微分为

$$\partial_B\phi(0,0)=\{(\xi-1,\eta-1)|\xi^2+\eta^2=1\}. \tag{8.5.6}$$

命题 8.5.2　Fischer-Burmeister NCP 函数 $\phi(a,b)$ 是正则的, 即广义方向导数与经典方向导数相等.

证明　根据命题 8.5.1, 我们只需考虑原点, 即证明

$$\phi'(0,d)=\phi^\circ(0,d), \quad \forall d\in\mathbf{R}^2.$$

先考虑 ϕ 在 0 点的经典方向导数, 记 $d=(d_1,d_2)\in\mathbf{R}^2$, 其中 $d_1,d_2\in\mathbf{R}$, 则有

$$\begin{aligned}\Delta\phi &=\phi(0+\lambda d)-\phi(0)=\phi(\lambda d)\\ &=\phi(\lambda d_1,\lambda d_2)=\sqrt{\lambda^2 d_1^2+\lambda^2 d_2^2}-\lambda d_1-\lambda d_2, \quad \lambda\geqslant 0,\end{aligned}$$

于是

$$\phi'(0,d)=\lim_{\lambda\to 0^+}\frac{\Delta\phi}{\lambda}=\sqrt{d_1^2+d_2^2}-d_1-d_2. \tag{8.5.7}$$

注意到

$$\sqrt{d_1^2+d_2^2}=\max_{u\in B((-1,-1),1)} u^{\mathrm{T}}d, \quad \forall\, \|d\|=1,$$

由式 (8.5.7) 得

$$\begin{aligned}\phi'(0,d) &= \max_{u\in B(0,1)} u^{\mathrm{T}}d - d_1 - d_2 \\ &= \max_{u\in B(0,1)} u^{\mathrm{T}}d - (1,1)d \\ &= \max_{u\in B(0,1)} u^{\mathrm{T}}d, \quad \forall \|d\| = 1.\end{aligned} \tag{8.5.8}$$

由于

$$\phi'((a,b),\lambda d) = \lambda\phi'((a,b),d), \quad \lambda \geqslant 0,$$

由式 (8.5.8) 得

$$\phi'(0,d) = \max_{u\in B((-1,-1),1)} u^{\mathrm{T}}d, \quad \forall d \in \mathbf{R}^2, \tag{8.5.9}$$

再由式 (8.5.6) 得

$$B((-1,-1),1) = \partial_B\phi(0). \tag{8.5.10}$$

结合式 (8.5.9), 式 (8.5.10) 以及广义方向导数定义及 $\mathrm{co}\partial_B\phi = \partial_{Cl}\phi$, 得

$$\phi'(0,d) = \max_{u\in B((-1,-1),1)} u^{\mathrm{T}}d = \phi^{\circ}(0,d), \quad \forall d \in \mathbf{R}^2,$$

$\phi(a,b)$ 在 0 点是正则的. 命题得证.

还有很多非线性互补函数, 这里不再介绍, 有兴趣的读者可参阅有关专著.

第 9 章　控制系统的生存性

本章利用非光滑分析理论研究控制系统的生存性, 在此假设读者对控制理论的基本理论和实际应用背景已有所了解, 有关内容不再做详细介绍.

9.1　微分包含与生存性

9.1.1　微分包含

除具有更一般的实际意义外, 微分包含形式表示的动态系统更方便于生存性问题研究, 因此我们以微分包含的概念来开始本章的内容. 考虑下述形式的微分包含:

$$\dot{x}(t) \in F(t, x), \tag{9.1.1}$$

其中 $F(t, x(t))$ 为 $\mathbf{R} \times \mathbf{R}^n$ 到 $\mathbf{R}^n$ 中子集的集值映射, 所谓微分包含 (9.1.1) 的解是指 $\mathbf{R}$ 到 $\mathbf{R}^n$ 上的绝对连续函数 $x(t)$, 它几乎处处满足式 (9.1.1).

通常的控制系统是微分包含的特殊形式, 当取

$$F(t, x) = \bigcup_{u \in U} f(t, x, u)$$

时, 微分包含 (9.1.1) 即为通常的非线性控制系统:

$$\dot{x}(t) \in f(t, x, u), \quad u \in U;$$

当取

$$F(t, x) = \bigcup_{u \in U} (Ax + Bu)$$

时, 微分包含 (9.1.1) 即为通常的线性控制系统:

$$\dot{x}(t) = Ax + Bu, \quad u \in U;$$

当取 $F(t, x)$ 为单点集, 即 $F(t, x) = \{f(t, x)\}$ 时, 微分包含 (9.1.1) 为通常的常微分方程:

$$\dot{x}(t) = f(t, x).$$

为保证微分包含 (9.1.1) 解的存在性以及有关性质, 通常对集值映射 $F(t, x)$ 做下述假设.

假设 9.1.1 集值映射 $F(t,x)$ 满足下述条件:

(1) 对任意 $(t,x) \in \mathbf{R} \times \mathbf{R}^n, F(t,x)$ 为 $\mathbf{R}^n$ 中非空凸紧集;

(2) 集值映射 $(t,x) \to F(t,x)$ 是上半连续的;

(3) 存在非负常数 γ, c, 使得对任意 (t,x), 有

$$v \in F(t,x) \Rightarrow ||v|| \leqslant \gamma ||v|| + c.$$

条件 (3) 称为线性增长条件, 它在微分包含和控制系统研究中有时也由 Lipschitz 条件代替, 事实上, Lipschitz 函数一定满足线性增长条件. 当 $F(t,x)$ 为单点集时, 记 $F(t,x) = \{f(t,x)\}$, 此时条件 (2) 等价于 $f(t,x)$ 的连续性, 条件 (3) 等价于

$$||f(t,x)|| \leqslant \gamma ||x|| + c.$$

9.1.2 生存性基本概念

生存性是控制理论中的一个重要领域, 控制理论中许多问题本质上都可以利用生存理论这一工具刻画并加以解决, 例如系统的可达性 (可控性)、Lyapunov 稳定性、微分对策等、另一方面, 系统的安全域设计本身就是一个生存性问题, 就是对给定系统设计一个生存域.

为简单起见, 考虑下述微分包含:

$$\dot{x}(t) \in F(x), \quad x \in \mathbf{R}^n, \tag{9.1.2}$$

这里 $F(x)$ 为 $\mathbf{R}^n$ 到 $\mathbf{R}^n$ 中子集的映射.

定义 9.1.1 设 $S \subset \mathbf{R}^n$, 如果对任意初始条件 $x_0 \in S$, 存在 (9.1.2) 的解 $x(t)$, 使得 $x(t) \in S$, $\forall t \geqslant 0$, 则称集合 S 关于微分包含 (9.1.2) 是生存的, 这样的解 $x(t)$ 也称为微分包含 (9.1.2) 的一个生存解.

假设 9.1.1 不但可以保证微分包含 (9.1.1) 和微分包含 (9.1.2) 解的存在, 还可以得到生存性的判别准则.

定理 9.1.1 如果假设 9.1.1 成立, 微分包含 (9.1.1) 在闭集 $S \subset \mathbf{R}^n$ 上是生存的充要条件是对任意 $x \in S$, 有

$$F(x) \bigcap T_S(x) \neq \varnothing, \tag{9.1.3}$$

其中 $\varnothing$ 代表空集.

对于集合 S 的内点 x, 易见 $T_S(x) = \mathbf{R}^n$, 因此, 对于集合 S 的内点, 式 (9.1.3) 总成立, 于是要判别式 (9.1.3) 是否成立, 只需考虑 S 的边界点.

考虑含有约束的微分包含:

$$\dot{x}(t) \in F(x), \quad x \in W, \tag{9.1.4}$$

这里 $F(x)$ 为 $\mathbf{R}^n$ 到 $\mathbf{R}^n$ 中子集的映射, W 为 $\mathbf{R}^n$ 中子集, 微分包含 (9.1.4) 在 W 中的最大生存域称为微分包 (9.1.4) 的生存核. 一般情况下, 生存核并不一定存在, 如果存在, 则是人们最感兴趣的一个生存域.

9.2 生存性的判别

定理 9.1.1 给出了一个闭集关于微分包含生存性的充要条件, 尽管在理论上是很完美的, 然而对一般的集合关于一般的微分包含或非线性控制系统, 验证式 (9.1.3) 是非常困难的, 甚至是不可能的. 本节讨论一些具体结构的集合关于特殊结构微分包含生存性的判别方法, 即对一个固定的 x, 判别 $F(x)\bigcap T_S(x)$ 是否非空.

9.2.1 不等式约束集合关于多面体微分包含

考虑微分包含:

$$\dot{x}(t) \in \mathrm{co}\{f_i(t)\,|\, i=1,\cdots,p\}\,,\quad x\in\mathbf{R}^n, \tag{9.2.1}$$

其中 $f_i(x)$, $i=1,\cdots,p$ 为 $\mathbf{R}^n$ 到 $\mathbf{R}^n$ 上的函数, 使得 $\mathrm{co}\{f_i(t)\,|\,i=1,\cdots,p\}$ 满足假设定理 9.1.1. 考虑下述形式的区域:

$$D=\{x\in\mathbf{R}^n\,|g_j(x)\leqslant 0,\ j=1,\cdots,m\}\,, \tag{9.2.2}$$

其中 $g_j(x), j=1,\cdots,m$ 为 $\mathbf{R}^n$ 上连续可微函数.

式 (9.1.3) 判别集合 D 关于微分包含 (9.2.1) 的生存性, 就是对每个给定的 $x\in\mathbf{R}^n$, 判别

$$\mathrm{co}\{f_i(x)\,|i=1,\cdots,p\}\bigcap T_D(x)\neq\varnothing \tag{9.2.3}$$

成立的方法.

给定 $x\in\mathbf{R}^n$, 定义指标集:

$$J(x)=\{j\in\{1,\cdots,m\}\,|g_j(x)=0\}\,.$$

如果 $x\in D$ 且 $J(x)$ 为空集, 则 x 为 D 的内点, 因此, 为判别式 (9.2.3) 是否成立, 只需考虑指标集 $J(x)$ 非空情况. 给出集合 D 在 x 点处的下述约束品性.

约束品性 9.2.1　存在 $y_0\in\mathbf{R}^n$, 使得

$$\nabla g_j(x)^{\mathrm{T}}y_0<0,\quad \forall j\in J(x).$$

约束品性 9.2.2　$\mathrm{cl}\,\gamma(x)=\Gamma(x)$, 其中

$$\gamma(x)=\{y\in\mathbf{R}^n\,\big|\nabla g_j(x)^{\mathrm{T}}y<0,\forall j\in J(x)\},$$

$$\Gamma(x)=\{y\in\mathbf{R}^n\,\big|\nabla g_j(x)^{\mathrm{T}}y\leqslant 0,\forall j\in J(x)\}.$$

根据非线性规则中约束品性的内容, 如果集合 D 在 $x\in\mathbf{R}^n$ 处满足约束品性 9.2.1 或约束品性 9.2.2, 则有 $T_D(x)=\Gamma(x)$, 也就是

$$T_D(x)=\{y\in\mathbf{R}^n\,\big|\nabla g_j(x)^{\mathrm{T}}y\leqslant 0,\forall j\in J(x)\}. \tag{9.2.4}$$

由式 (9.2.4) 易见, 切锥 $T_D(x)$ 是一个多面体, 因此, 可以利用它的这一特殊结构以及 $F(x)$ 的结构判别式 (9.2.3) 是否成立. 首先, 考虑两个多面体集合:

$$U=\mathrm{co}\{u_i\,|\,i=1,\cdots,p\},$$

$$V=\{y\in\mathbf{R}^n\,|My\leqslant 0\},$$

其中 $u_i\in\mathbf{R}^n$, M 是 $m\times n$ 阶矩阵. 易见, 集合 $\mathrm{co}\{f_i(x)\,|i=1,\cdots,p\}$ 和集合 $T_D(x)$ 可分别表示为 U 和 V 的形式. 构造下述线性不等式组:

$$\sum_{i=1}^{p}\lambda_iMu_i\leqslant 0,\quad \sum_{i=1}^{p}\lambda_i=1,\quad \lambda_i\geqslant 0,\quad i=1,\cdots,p, \tag{9.2.5}$$

其中 $(\lambda_1,\cdots,\lambda_p)$ 为变量. 易见, 不等式组 (9.2.5) 有 $m+p$ 个不等式和 1 个等式.

定理 9.2.1　集合 $U\bigcap V$ 非空当且仅当不等式组 (9.2.5) 是相容 (有解) 的.

证明　必要性. 假设 $U\bigcap V$ 非空, 记 $z\in U\bigcap V$, 于是 $z\in U$, $z\in V$. 根据凸包的定义, 存在 $\lambda_i\geqslant 0$, $i=1,\cdots,p$ 满足

$$\sum_{i=1}^{p}\lambda_i=1,$$

使得

$$z=\sum_{i=1}^{p}\lambda_iu_i.$$

另一方面, $z\in V$ 表明 $Mz\leqslant 0$, 于是

$$Mz=M\sum_{i=1}^{p}\lambda_iu_i=\sum_{i=1}^{p}\lambda_iMu_i\leqslant 0.$$

以上讨论表明 $\lambda_i\geqslant 0$, $i=1,\cdots,p$ 是不等式组 (9.2.5) 的解, 于是不等式组 (9.2.5) 是相容的.

充分性. 假设不等式组 (9.2.5) 是相容的, 记 $(\lambda_1^*,\cdots,\lambda_p^*)$ 是它的一个解, 于是有

$$\sum_{i=1}^{p}\lambda_i^*Mu_i\leqslant 0,\quad \sum_{i=1}^{p}\lambda_i^*=1,\quad \lambda_i^*\geqslant 0,\quad i=1,\cdots,p. \tag{9.2.6}$$

令

$$z^* = \sum_{i=1}^{p} \lambda_i^* u_i,$$

根据式 (9.2.6), 易见 $z^* \in U$. 另一方面, 由式 (9.2.6) 可知

$$Mz^* = \sum_{i=1}^{p} \lambda_i^* M u_i \leqslant 0,$$

于是 $z^* \in V$. 综上可得 $z^* \in U \bigcap V$, 集合 $U \bigcap V$ 非空. 定理得证.

命题 9.2.1 如果存在指标 $1 \leqslant i_0 \leqslant p$, 使得 $Mu_{i_0} \leqslant 0$, 则 $u_{i_0} \in U \bigcap V$, 即 $U \bigcap V$ 非空.

证明 选取

$$\lambda_{i_0}^* = 1, \quad \lambda_k^* = 0, \quad k \in \{1, \cdots, p\} \backslash \{i_0\},$$

可以验证 $(\lambda_1^*, \cdots, \lambda_p^*)$ 是不等式组 (9.2.5) 的一个解, 根据定理 9.2.1, $U \bigcap V$ 非空. 命题得证.

基于定理 9.2.1, 判别集合 $U \bigcap V$ 非空性可转化为判别线性不等式组 (9.2.6) 的相容性. 事实上, 判别线性不等式组的相容性可转化为求解一个辅助线性规划问题, 有下述事实.

设 A 是 $s \times n$ 阶矩阵, B 是 $t \times n$ 阶矩阵, 则线性不等式组

$$Ay \leqslant 0, \quad By = b, \quad y \in \mathbf{R}^n$$

相容 (有解) 的充要条件是线性规划问题:

$$\begin{aligned} &\min w, \\ &\text{s.t. } Ay - (w, \cdots, w)^{\mathrm{T}} \leqslant 0, \\ &\qquad By = b, \quad w \geqslant 0 \end{aligned}$$

的最优值为零, 其中 $y \in \mathbf{R}^n$, $w \in \mathbf{R}$ 为变量, $(w, \cdots, w) \in \mathbf{R}^m$.

基于定理 9.1.1、定理 9.2.1、命题 9.2.1 和式 (9.2.4), 在约束品性 9.2.1 或约束品性 9.2.2 成立的条件下, 给出判别式 (9.2.3) 在 x 点处是否成立 (是否满足生存性条件) 的算法.

算法 9.2.1

步 1. 给定 $x \in D$, 计算函数 $f_i(x), i = 1, \cdots, p$ 和 $g_i(x), j = 1, \cdots, m$ 在点 x 的值, 确定指标集 $J(x)$, 记 $J(x) = \{j_1, \cdots, j_t\}$, 计算梯度值 $\nabla g_j(x), j \in J(x)$;

步 2. 计算向量:

$$(\nabla g_{j_1}^{\mathrm{T}}(x), \cdots, \nabla g_{j_t}^{\mathrm{T}}(x))^{\mathrm{T}} f_i(x), \quad i = 1, \cdots, p,$$

如果存在指标 $1 \leqslant i_0 \leqslant p$, 使得

$$(\nabla g_{j_1}^{\mathrm{T}}(x), \cdots, \nabla g_{j_t}^{\mathrm{T}}(x))^{\mathrm{T}} f_{i_0}(x) \leqslant 0,$$

则

$$\mathrm{co}\,\{f_i(x)\,|i=1,\cdots,p\} \bigcap T_D(x) \neq \varnothing,$$

停止, 否则转步 3;

步 3. 解线性规划问题:

$$\begin{aligned}
&\min w,\\
&\text{s.t. } \sum_{i=1}^{p} \lambda_i (\nabla g_{j_1}^{\mathrm{T}}(x), \cdots, \nabla g_{j_t}^{\mathrm{T}}(x))^{\mathrm{T}} f_i(x) + (w, \cdots, w)^{\mathrm{T}} \leqslant 0,\\
&\quad\ \sum_{i=1}^{p} \lambda_i = 1, w - \lambda_i \geqslant 0, i = 1, \cdots, p,
\end{aligned} \tag{9.2.7}$$

如果问题 (9.2.7) 的最优值为零, 则有

$$\mathrm{co}\,\{f_i(x)\,|i=1,\cdots,p\} \bigcap T_D(x) \neq \varnothing,$$

否则

$$\mathrm{co}\,\{f_i(x)\,|i=1,\cdots,p\} \bigcap T_D(x) = \varnothing.$$

9.2.2 非光滑边界区域关于多面体微分包含

考虑微分包含:

$$\dot{x}(t) \in \mathrm{co}\{f_i(x)\,|\,i=1,\cdots,p\}\,, \quad x \in \mathbf{R}^n \tag{9.2.8}$$

和区域:

$$W = \{x \in \mathbf{R}^n\,|g_j(x) \leqslant 0,\ j=1,\cdots,m\}\,, \tag{9.2.9}$$

其中 $f_i(x)$, $i=1,\cdots,p$ 为 $\mathbf{R}^n$ 到 $\mathbf{R}^n$ 上的适当函数, 使得 $\mathrm{co}\{f_i(t)\,|\,i=1,\cdots,p\}$ 满足假设 9.1.1, $g_j(x)$, $j=1,\cdots,m$ 为 $\mathbf{R}^n$ 上的次可微函数.

由第 5 章内容知, 如果 $h_i(x), i \in I$ 为次可微函数, 则

$$h(x) = \max_{i \in I} h_i(x)$$

亦为次可微函数, 其次微分为

$$\partial h(x) = \mathrm{co} \bigcup_{i \in I(x)} \partial h_i(x),$$

其中

$$I(x) = \{\, i \in I | h_i(x) = h(x)\,\}.$$

由此易见, 如果每个次微分 $\partial h_i(x)$ 为有限点集的凸包, 则 $\partial h(x)$ 也为有限点集凸包. 以下假设每个次可微函数 $g_j(x)$ 的次微分 $\partial g_j(x)$ 为有限点集凸包. 令

$$g(x) = \max_{1 \leqslant j \leqslant m} g_j(x),$$

由于

$$\max_{1 \leqslant j \leqslant m} g_j(x) \leqslant 0$$

等价于

$$g_j(x) \leqslant 0, \quad j = 1, \cdots, m,$$

于是集合 (9.2.9) 可等价地表示为

$$W = \{x \in \mathbf{R}^n | g(x) \leqslant 0\}.$$

由于 $g_j(x)$ 是次可微的, 则 $g(x)$ 也是次可微的, 又 $g_j(x)$ 的次微分为有限点集凸包, 故 $g(x)$ 的次微分也为有限点集凸包. 记

$$\partial g(x) = \mathrm{co}\{v_1, \cdots, v_q\}, \tag{9.2.10}$$

其中 $v_i \in \mathbf{R}^n$, $i = 1, \cdots, q$, 定义矩阵

$$B = (v_1, \cdots, v_q)^{\mathrm{T}}. \tag{9.2.11}$$

为判别微分包含 (9.2.8) 在集合 (9.2.9) 上的生存性条件, 对于给定的 $x \in \mathbf{R}^n$, 要判别下式是否成立:

$$\mathrm{co}\{f_i(x) \,|i = 1, \cdots, p\} \bigcap T_D(x) \neq \varnothing. \tag{9.2.12}$$

为此, 先给出下面两个约束品性, 它们在非光滑优化中也是经常使用的.

约束品性 9.2.3　存在 $y_0 \in \mathbf{R}^n$, 使得 $g'(x; y_0) < 0$.

约束品性 9.2.4　$\mathrm{cl}\,\gamma(x) = \Gamma(x)$ 成立, 其中

$$\gamma(x) = \{y \in \mathbf{R}^n \,|g'(x; y) < 0\},$$

$$\Gamma(x) = \{y \in \mathbf{R}^n \,|g'(x; y) \leqslant 0\}.$$

定理 9.2.2　假设约束品性 9.2.3 或约束品性 9.2.4 成立, 则有 $T_D(x) = \{y \in \mathbf{R}^n | By \leqslant 0\}$.

证明　根据非光滑优化约束条件有关知识, 当约束品性 9.2.3 或约束品性 9.2.4 成立时, 则有 $T_D(x) = \Gamma(x)$, 因此, 只需证明对固定的 $x \in \mathbf{R}^n$, $g'(x; y) \leqslant 0$ 等价于 $By \leqslant 0$. 注意到,

$$g'(x; y) = \max_{v \in \partial g(x)} v^{\mathrm{T}} y = \max_{1 \leqslant i \leqslant q} v_i^{\mathrm{T}} y, \tag{9.2.13}$$

而 $\max\limits_{1\leqslant i\leqslant q} v_i^{\mathrm{T}}y \leqslant 0$ 等价于

$$v_i^{\mathrm{T}}y \leqslant 0, \quad \forall\, 1 \leqslant i \leqslant q,$$

根据矩阵 B 的定义,

$$v_i^{\mathrm{T}}y \leqslant 0, \quad \forall\, 1 \leqslant i \leqslant m$$

等价于 $By \leqslant 0$, 再由式 (9.2.13) 得 $g'(x;y) \leqslant 0$ 等价于 $By \leqslant 0$. 定理得证.

根据定理 9.2.2, 判别式 (9.2.12) 可转化为判别下式:

$$\mathrm{co}\{f_i(x)|i = 1,\cdots,p\} \bigcap \{y \in \mathbf{R}^n|By \leqslant 0\} \neq \varnothing. \tag{9.2.14}$$

构造下述线性不等式组:

$$\sum_{i=1}^{p} \lambda_i B f_i(x) \leqslant 0, \quad \sum_{i=1}^{p} \lambda_i = 1, \quad \lambda_i \geqslant 0, \quad i = 1,\cdots,p, \tag{9.2.15}$$

其中 $(\lambda_1,\cdots,\lambda_p) \in \mathbf{R}^p$ 为变量, x 为 $\mathbf{R}^n$ 中固定点. 显然, 不等式组 (9.2.15) 有 $m+p$ 个不等式和 1 个等式.

类似本节前一部分的讨论知, 式 (9.2.14) 成立的充要条件为线性不等式组 (9.2.15) 有解. 这样, 可通过判别线性不等式组 (9.2.15) 是否有解来检验生存性条件 (9.2.12) 是否成立, 当然, 也可将判别线性不等式组是否有解等价地转化为求解一个线性规划问题, 因此是非常容易实现的.

例 9.2.1 在微分包含式 (9.2.8) 和式 (9.2.9) 给出的集合 W 中, 设

$$f_1(x) = (x_1, x_1 - x_2)^{\mathrm{T}}, \quad f_2(x) = (x_1 + x_2, x_2)^{\mathrm{T}},$$

$$g(x) = \max\left\{-x_1, -x_2, x_1^2 + x_2^2 - 1\right\}, \quad x = (x_1, x_2)^{\mathrm{T}},$$

其中 $x = (x_1, x_2)^{\mathrm{T}} \in \mathbf{R}^2$. 易见 $g(x)$ 为次可微函数, 集合 $W = \{x \in \mathbf{R}^2|g(x) \leqslant 0\}$ 为四分之一单位圆. 考虑点 $x^{(1)} = (0,1)^{\mathrm{T}}$ 和 $x^{(2)} = (1,0)^{\mathrm{T}}$, 根据第 5 章介绍的次微分定义及运算得

$$\partial g(x^{(1)}) = \mathrm{co}\left\{(-1,0)^{\mathrm{T}}, (0,2)^{\mathrm{T}}\right\}, \quad \partial g(x^{(2)}) = \mathrm{co}\left\{(0,-1)^{\mathrm{T}}, (2,0)^{\mathrm{T}}\right\}.$$

$f_i(x)$ 在 $x^{(1)}$ 和 $x^{(2)}$ 处的值为

$$f_1(x^{(1)}) = (0,-1)^{\mathrm{T}}, \quad f_2(x^{(1)}) = (1,1)^{\mathrm{T}},$$

$$f_1(x^{(2)}) = (1,1)^{\mathrm{T}}, \quad f_2(x^{(2)}) = (1,0)^{\mathrm{T}}.$$

通过验证可知, 微分包含在 $x^{(1)} = (0,1)^{\mathrm{T}}$ 处满足生存性条件 (9.2.12), 在 $x^{(2)} = (1,0)^{\mathrm{T}}$ 处不满足生存性条件 (9.2.12).

下面讨论次可微函数上图的生存性判别. 设 $V(x)$ 为 $\mathbf{R}^n$ 上的次可微函数, 它的上图为

$$\mathrm{Epi}(V) = \{(x,w) \in \mathbf{R}^{n+1} | V(x) - w \leqslant 0\},$$

令 $H(z) = V(x) - w$, 其中 $z = (x,w)$, 则有

$$\mathrm{Epi}(V) = \{z \in \mathbf{R}^{n+1} | H(z) \leqslant 0\}.$$

记

$$S(z) = \{(u,-1) | u \in \partial V(x)\},$$

通过计算得

$$H'(z;d) = \max_{\xi \in S(z)} \xi^{\mathrm{T}} d,$$

根据次可微函数定义, $H(z)$ 是次可微的其次微分为

$$\partial H(z) = S(z) = \{(u,-1) | u \in \partial V(x)\}.$$

假设 $V(x)$ 的次微分为有限点集的凸包, 记为式 (9.2.10) 的形式, 则

$$\partial H(x) = \mathrm{co}\{(v_1,-1),\cdots,(v_q,-1)\}$$

也为有限点集凸包, 这样, 前面讨论的方法可用来判别上图 $\mathrm{Epi}(V)$ 的生存性问题.

在稳定性和镇定设计中, 经常遇到的非光滑 Lyapunov 函数是次可微的, 且次微为有限点集凸包. 例如, Lyapunov 函数 $V(x) = \max\limits_{i \in I} V_i(x)$, 其中 $V_i(x)$, $i \in I$ 为连续可微函数, 是次可微的, 其次微分为有限点集凸包. 这样, 可以通过 Lyapunov 函数上图的生存性研究 Lyapunov 稳定性和镇定设计.

9.2.3 一维生存核的计算

前面讨论的是判别一个给定集合的生存性, 一般来讲, 计算一个控制系统的生存域则更加困难. 一维系统具有极其特殊性, 在一维空间中, 凸集就是一个闭区间, 因此有可能计算一维控制系统的一个生存域.

在一维情形下考虑微分包含式 (9.1.2), 对任意 $x \in \mathbf{R}$, 由于集合 $F(x)$ 是闭凸的, 这样可表示为 $F(x) = [f_1(x), f_2(x)]$, 其中 $f_1(x)$ 和 $f_2(x)$ 为 $\mathbf{R}$ 上函数. 给定 $\mathbf{R}$ 中闭区间 $W_0 = [w_1, w_2]$, 以下讨论计算 W_0 中生存核. 设 $W = [z_1, z_2]$ 为微分包含在 W_0 中的生存核, 其中 z_1 和 z_2 待定. 因为只有 z_1 和 z_2 是 $W = [z_1, z_2]$ 的边界点, 式 (9.1.3) 成立的充要条件是其在 $x = z_1$ 和 $x = z_2$ 两点处成立, 即

$$F(z_1) \bigcap T_W(z_1) \neq \varnothing \quad 和 \quad F(z_2) \bigcap T_W(z_2) \neq \varnothing.$$

由切锥的定义得

$$T_W(z_1) = [0, +\infty), \quad T_W(z_2) = (-\infty, 0],$$

于是有

$$F(z_1)\bigcap T_W(z_1)\neq\varnothing\Leftrightarrow f_2(z_1)\geqslant 0, \tag{9.2.16}$$

$$F(z_2)\bigcap T_W(z_2)\neq\varnothing\Leftrightarrow f_1(z_2)\leqslant 0. \tag{9.2.17}$$

满足式 (9.2.16) 和式 (9.2.17) 的区间 $[z_1,z_2]\subset W_0$ 即为微分包含的生存核.

为确定生存核 $W=[z_1,z_2]$, 考虑下述两个优化问题:

$$\begin{aligned}&\min z,\\ &\text{s.t. } f_2(z)\geqslant 0,\quad w_1\leqslant z\leqslant w_2,\end{aligned} \tag{9.2.18}$$

$$\begin{aligned}&\max z,\\ &\text{s.t. } f_1(z)\leqslant 0,\quad w_1\leqslant z\leqslant w_2.\end{aligned} \tag{9.2.19}$$

设 z_1 和 z_2 分别是问题 (9.2.18) 和 (9.2.19) 的最优解, 由式 (9.2.16) 和式 (9.2.17) 可知, 如果 $z_1\leqslant z_2$, 则 $W=[z_1,z_2]$ 是生存域, 且是生存域核, 否则生存域是空集. 显然, 问题 (9.2.18) 和问题 (9.2.19) 是一维约束优化问题, 一般情况下是非光滑优化问题.

考虑非线性控制系统:

$$\dot{x}(t)=f(x,u),\quad u\in U, \tag{9.2.20}$$

其中 $x\in\mathbf{R},U\subset\mathbf{R}$ 为紧集. 令

$$F(x)=\{f(x,u)\mid u\in U\},$$

将系统 (9.2.20) 看成微分包含 $\dot{x}(t)\in F(x)$ 的特殊情形. 由于 $F(x)$ 为闭凸集, 它可表示为

$$F(x)=[\min_{u\in U}f(x,u),\max_{u\in U}f(x,u)],$$

即

$$f_1(x)=\min_{u\in U}f(x,u)\quad 和\quad f_2(x)=\max_{u\in U}f(x,u).$$

对于系统 (9.2.20), 问题 (9.2.18) 和问题 (9.2.19) 具有形式:

$$\begin{aligned}&\min z,\\ &\text{s.t. } \max_{u\in U}f(z,u)\geqslant 0,\quad w_1\leqslant z\leqslant w_2,\end{aligned} \tag{9.2.21}$$

$$\begin{aligned}&\max z,\\ &\text{s.t. } \max_{u\in U}f(z,u)\geqslant 0,\quad w_1\leqslant z\leqslant w_2.\end{aligned} \tag{9.2.22}$$

易见, 问题 (9.2.21) 和 (9.2.22) 是一维非光滑约束优化问题, 即使 $f(z,\ u)$ 是连续可微的也是如此.

综上所述, 可以通过求解一维非光滑约束优化问题来计算一维微分包含生存核.

例 9.2.2　设

$$F(x)=[f_1(x),f_2(x)],\quad f_1(x)=x^2+x,\quad f_2(x)=2x^2+x,$$

求微分包含 $\dot{x}(t)\in F(x), x\in\mathbf{R}$ 在集合 $W_0=[-1,1]$ 中的生存核.

将 $f_2(x)=2x^2+x$ 和 $W_0=[-1,\,1]$ 代入优化问题 (9.2.18) 得

$$\begin{aligned}&\min z,\\&\text{s.t. }\ 2z^2+z\geqslant 0,\quad -1\leqslant z\leqslant 1,\end{aligned}$$

其最优解为 $z_1=-1$. 将 $f_1(x)=x^2+x$ 和 $W_0=[-1,1]$ 代入优化问题 (9.2.19) 中得

$$\begin{aligned}&\max z,\\&\text{s.t. }\ 2z^2+z\geqslant 0,\quad -1\leqslant z\leqslant 1,\end{aligned}$$

其最优解为 $z_2=0$. 于是, 生存核为 $W=[-1,0]$.

例 9.2.3　设 $f(x)=x^2+u$, 其中 $u\in U=[0,1]$, $x\in\mathbf{R}$, 求非线性系统 $\dot{x}(t)=f(x,u)$ 在集合 $W_0=[-1,1]$ 中的生存核.

由于

$$\max_{u\in U}f(x)=x^2+1\quad 和\quad \min_{u\in U}f(x)=x^2,$$

将

$$\max_{u\in U}f(x)=x^2+1\quad 和\quad W_0=[-1,1]$$

代入问题 (9.2.21) 中得

$$\begin{aligned}&\min\ \ z,\\&\text{s.t. }\ z^2+1\geqslant 0,\quad -1\leqslant z\leqslant 1,\end{aligned}$$

其最优解为 $z_1=-1$. 将

$$\min_{u\in U}f(x)=x^2\quad 和\quad W_0=[-1,\,1]$$

代入优化问题 (9.2.22) 得

$$\begin{aligned}&\max\ \ z,\\&\text{s.t. }\ z^2\leqslant 0,\quad -1\leqslant z\leqslant 1,\end{aligned}$$

其最优解为 $z_2=0$. 于是, 生存域核为 $[-1;0]$.

9.3 线性系统多面体生存域

线性系统的生存性是生存性理论中的重要内容, 讨论其生存域主要有两种方法: 一种是利用矩阵不等式或半定规划得到一个椭球生存域; 另一种就是利用凸分析方法研究的多面体生存域. 由于多面体可以逼近任意的区域, 因此多面体生存域比较椭球生存域来讲, 更有应用价值.

9.3.1 有界生存域

考虑带有状态约束的线性控制系统:

$$\dot{x}(t) = Ax + Bu, \quad x \in K, \quad u \in U, \tag{9.3.1}$$

其中 $K \subset \mathbf{R}^n$, $U \subset \mathbf{R}^m$ 为闭凸集, A 为适当维数的矩阵.

定理 9.3.1 如果 $D \subset \mathbf{R}^n$ 是系统 (9.3.1) 的生存域, 则它的凸包 $\text{co}D$ 也为系统 (9.3.1) 的生存域, 进一步生存核为凸集.

证明 注意到系统 (9.3.1) 的解为

$$x(t) = \mathrm{e}^{At}x_0 + \int_o^t \mathrm{e}^{A(t-\tau)}Bu(\tau)\mathrm{d}\tau,$$

其中 $x(0) = x_0$. 给定 $x_0 \in \text{co}D$, 根据第 1 章凸集的性质, 存在 $x_i \in D$, $i = 1, \cdots, n+1$, 使得 x_0 可表示为 $x_i\ i = 1, \cdots, n+1$ 的凸组合, 即

$$x_0 = \sum_{i=1}^{n+1} \lambda_i x_i, \quad \lambda_i \geqslant 0, \quad i = 1, \cdots, n+1, \quad \sum_{i=1}^{n+1} \lambda_i = 1.$$

由于 D 是生存域, 而 $x_i \in D$, 因此, 存在 $u_i(\cdot)$:$[0, \infty) \to U$, 使得对任意 $t > 0$, 有

$$\mathrm{e}^{At}x_i + \int_o^t \mathrm{e}^{A(t-\tau)}Bu_i(\tau)\mathrm{d}\tau \in D.$$

令

$$x(t) = \sum_{i=1}^{n+1} \lambda_i \left(\mathrm{e}^{At}x_i + \int_o^t \mathrm{e}^{A(t-\tau)}Bu_i(\tau)\mathrm{d}\tau\right),$$

显然 $x(t) \in \text{co}D$. 令

$$u(t) = \sum_{i=1}^{n+1} \lambda_i u_i(t), \quad \forall t \in [0, \infty),$$

由于 U 是凸集, 因此

$$u(t) \in U, \quad \forall t \geqslant 0.$$

此外

$$x(t) = e^{At}\sum_{i=1}^{n+1}\lambda_i x_i + \int_0^t e^{A(t-\tau)}B\sum_{i=1}^{n+1}\lambda_i u_i(\tau)d\tau$$
$$= e^{At}x_0 + \int_o^t e^{A(t-\tau)}Bu(\tau)d\tau,$$

所以 $x(t)$ 是线性系统 (9.3.1) 当初值条件为 x_0 的解, 且使得 $x(t) \in \mathrm{co}D, \forall t \geqslant 0$, 因此 $\mathrm{co}D$ 是生存域.

由于生存核是最大的生存域, 所以它与自身的凸包相等, 故生存核是凸集. 定理得证.

定理 9.3.2 设 $W = \mathrm{co}\{w_1, \cdots, w_l\}$, 其中 $w_i \in \mathbf{R}^n$, $i = 1, \cdots, l$, $U = \mathrm{co}\{\hat{u}_1, \cdots, \hat{u}_M\}$, 其中 $\hat{u}_j \in \mathbf{R}^m, j = 1, \cdots, M$, 则 W 是系统 (9.3.1) 的生存域当且仅当

$$T_W(w_i) \bigcap F(w_i) \neq \varnothing, \quad i = 1, \cdots, l,$$

其中 $F(x) = \{Ax + Bu | u \in U\}$.

证明 将系统 (9.3.1) 表示为微分包含 $\dot{x}(t) \in F(x)$ 的形式. 根据定理 9.1.1, 必要性成立, 现在证明充分性. 给定 $x \in W$ 和某个 $y \in F(x)$, 则存在

$$\mu_j \geqslant 0, \quad j = 1, \cdots, M,$$

满足

$$\sum_{j=1}^{M}\mu_j = 1,$$

使得

$$y = Ax + \sum_{j=1}^{M}\mu_j B\hat{u}_j.$$

对于多面体 W, 根据切锥定义, $y \in T_W(x)$ 当且仅当存在 $t > 0$ 和 $v_j \geqslant 0, j = 1, \cdots, l$ 满足

$$\sum_{j=1}^{l} v_j = 1,$$

使得

$$y = \frac{1}{t}\left(\sum_{j=1}^{l} v_j w_j - x\right).$$

所以, $T_W(x) \bigcap F(x) \neq \varnothing$ 当且仅当对某个 $t > 0$, 下式成立:

$$\left(A + \frac{1}{t}I\right)x = \frac{1}{t}\sum_{j=1}^{l} v_j w_j - \sum_{j=1}^{M}\mu_j B\hat{u}_j, \tag{9.3.2}$$

其中

$$v_j \geqslant 0, \quad j=1,\cdots,l, \quad \sum_{j=1}^{l} v_j = 1, \quad \mu_j \geqslant 0, \quad j=1,\cdots,M \text{ 且 } \sum_{j=1}^{l}\mu_j = 1.$$

假设对于 $x = w_i, i=1,\cdots,l$, 式 (9.3.2) 成立, 则存在

$$v_{ij} \geqslant 0, \quad j=1,\cdots,l, \quad \sum_{j=1}^{l} v_{ij} = 1, \quad \mu_{ij} \geqslant 0, \quad j=1,\cdots,M \text{ 且 } \sum_{j=1}^{l}\mu_{ij} = 1,$$

使得

$$\left(A+\frac{1}{t_i}I\right)w_i = \frac{1}{t_i}\sum_{j=1}^{l} v_{ij}w_j - \sum_{j=1}^{M}\mu_{ij}B\hat{u}_j. \tag{9.3.3}$$

因为 W 是多面体, 所以 $x+\bar{t}y \in W$ 意味着

$$x+ty \in W, \quad \forall t \leqslant \bar{t}.$$

令 $t = \min\{t_1,\cdots,t_l\}$, 那么由式 (9.3.3) 可得

$$\left(A+\frac{1}{t}I\right)w_i = \frac{1}{t}\sum_{j=1}^{l} v_{ij}w_j - \sum_{j=1}^{M}\mu_{ij}B\hat{u}_j. \tag{9.3.4}$$

对于 $x \in W$, 根据集合 W 的结构, 存在 $\lambda_i \geqslant 0, i=1,\cdots,l$ 满足

$$\sum_{j=1}^{l}\lambda_j = 1,$$

使得

$$x = \sum_{i=1}^{l}\lambda_i w_i.$$

将式 (9.3.3) 关于 $\lambda_i, i=1,\cdots,l$ 取加权平均, 再利用

$$x = \sum_{i=1}^{l}\lambda_i w_i$$

得

$$\begin{aligned}\left(A+\frac{1}{t}I\right)x &= \frac{1}{t}\sum_{i=1}^{l}\lambda_i\sum_{j=1}^{l} v_{ij}w_j - \sum_{i=1}^{l}\lambda_i\sum_{j=1}^{M}\mu_{ij}B\hat{u}_j \\ &= \frac{1}{t}\sum_{j=1}^{l} v_j w_j - \sum_{j=1}^{M}\mu_j B\hat{u}_j,\end{aligned}$$

其中

$$v_j = \sum_{i=1}^{l} \lambda_i v_{ij}, \quad \mu_j = \sum_{i=1}^{l} \lambda_i \mu_{ij}.$$

注意到

$$v_j \geqslant 0, \quad j = 1, \cdots, l, \quad \mu_j \geqslant 0, \quad j = 1, \cdots, M,$$

且有

$$\sum_{j=1}^{M} \mu_j = \sum_{j=1}^{M} \sum_{i=1}^{l} \lambda_i \mu_{ij} = \sum_{i=1}^{l} \lambda_i \sum_{j=1}^{M} \mu_{ij} = 1.$$

同理可证

$$\sum_{j=1}^{l} v_j = 1.$$

这说明式 (9.3.2) 对任意的 x 都成立, 于是

$$T_W(x) \bigcap F(x) \neq \varnothing.$$

定理得证.

由定理 9.3.2 的证明可知, 多面体 W 是生存域当且仅当对每个 $i \in \{1, \cdots, l\}$, 存在足够小的 $t_i > 0$, 使得线性不等式系统:

$$\begin{cases} \left(A + \dfrac{1}{t_i} I\right) w_i = \dfrac{1}{t_i} \displaystyle\sum_{j=1}^{l} v_{ij} w_j - \sum_{j=1}^{M} \mu_{ij} B \hat{u}_j, \\ \displaystyle\sum_{j=1}^{l} v_{ij} = 1, \quad \sum_{j=1}^{M} \mu_{ij} = 1, \\ v_{ij} \geqslant 0, \quad j = 1, \cdots, l, \quad \mu_{ij} \geqslant 0, \quad j = 1, \cdots, M \end{cases} \tag{9.3.5}$$

是相容的. 确定不等式组 (9.3.2) 的相容性可转化为解一个辅助的线性规划问题, 因而是很容易实现的.

定理 9.3.2 说明, 对于有界多面体 (有限点集的凸包), 其生存性判别只需检验其在极点处是否满足生存性条件, 因而是简便易行的.

9.3.2　无界生存域

9.3.1 节讨论的是有界多面体的生存域, 本节讨论无界多面体情形. 根据 1.4 节知, 多面体结构可以表示为它的极点凸组合加上极方向的非负线性组合, 因此, 关键所在是讨论或确定生存域的极方向.

定理 9.3.3　设 $W \subset \mathbf{R}^n$ 为闭集, $\alpha \in \mathbf{R}^n$, 则对任意实数 s, $t \geqslant 0$, 有

$$T_W(x) + s\alpha \subset T_{W+\mathrm{cone}\{\alpha\}}(x + t\alpha).$$

证明 设 $u \in T_W(x)$, 根据切锥性质, 存在 $t_k > 0$, $u_k \in \mathbf{R}^n, k = 1, 2, \cdots$, 满足

$$u_k \to u, \quad t_k \to 0, \quad k \to \infty,$$

使得

$$x + t_k u_k \in W, \quad \forall k \geqslant 1.$$

对任意固定的 s, $t \geqslant 0$, 有

$$\begin{aligned} u_k + s\alpha &\to u + s\alpha, \\ x + t\alpha + t_k(u_k + s\alpha) &= x + t_k u_k + (s+t)\alpha \\ &\in W + \text{cone}\{\alpha\}. \end{aligned}$$

于是, 根据切锥性质得

$$u + s\alpha \in T_{W+\text{cone}\{\xi\}}(x + t\alpha),$$

再由 $u \in T_W(x)$ 的任意性, 则有

$$T_W(x) + s\alpha \subset T_{W+\text{cone}\{\alpha\}}(x + t\alpha).$$

定理得证.

定理 9.3.4 设 ξ 为 A 的具有非负特征值的特征向量, W 为系统 (9.3.1) 的生存域, 则集合 $W + \text{cone}\{\xi\}$ 也为系统 (9.3.1) 的生存域.

证明 设 $y \in W + \text{cone}\{\xi\}$, 则 y 可表示为 $y = x + t\xi$, 其中 $x \in W$, $t \geqslant 0$. W 是生存域, 根据定理 9.1.1, 有

$$T_W(x) \bigcap \{Ax + Bu \mid u \in U\} \neq \varnothing,$$

于是存在 $u_1 \in U$, 使得 $Ax + Bu_1 \in T_W(x)$. 记 ξ 的特征值为 $\lambda \geqslant 0$, 则有

$$\begin{aligned} Ay + Bu_1 &= A(x + t\xi) + Bu_1 \\ &= Ax + Bu_1 + t\lambda\xi. \end{aligned} \tag{9.3.6}$$

由于 $Ay + Bu_1 \in T_W(x)$, 以及 $t \geqslant 0$, $t\lambda \geqslant 0$, 根据定理 9.3.3 得

$$\begin{aligned} Ax + Bu_1 + t\lambda\xi &\in T_{W+\text{cone}\{\xi\}}(x + t\xi) \\ &= T_{W+\text{cone}\{\xi\}}(y). \end{aligned} \tag{9.3.7}$$

由式 (9.3.6) 和式 (9.3.7) 可得

$$Ay + Bu_1 \in T_{W+\text{cone}\{\xi\}}(y),$$

于是

$$\{Ay+Bu \mid u\in U\}\bigcap T_{W+\mathrm{cone}\{\xi\}}(y)\neq\varnothing.$$

由定理 9.1.1, $W+\mathrm{cone}\{y\}$ 是生存的. 定理得证.

推论 9.3.1　设 $\xi_1,\cdots,\ \xi_p$ 为 A 的具有非负特征值的特征向量, W 为系统 (9.3.1) 的生存域, 则集合 $W+\mathrm{cone}\{\xi_1,\ \cdots,\ \xi_p\}$ 也为系统 (9.3.1) 的生存域.

证明　根据凸锥的定义, 有

$$W+\mathrm{cone}\{\xi_1,\ \cdots,\ \xi_p\}=W+\mathrm{cone}\{\xi_1,\ \cdots,\ \xi_{p-1}\}+\mathrm{cone}\{\xi_p\}, \tag{9.3.8}$$

根据定理 9.3.4 和式 (9.3.8), 利用数学归纳法即得推论. 推论得证.

例 9.3.1　考虑线性系统:

$$\begin{pmatrix}\dot{x}_1(t)\\ \dot{x}_2(t)\end{pmatrix}=\begin{pmatrix}-1 & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix}. \tag{9.3.9}$$

可以验证单位圆

$$B(0,1)=\{(x_1,\ x_2)^{\mathrm{T}}\in\mathbf{R}^2\,\big|x_1^2+x_2^2\leqslant 1\}$$

是系统 (9.3.9) 的一个生存域, $\xi=(0,\ 1)^{\mathrm{T}}$ 是系数矩阵特征值为 1 的特征向量. 由定理 9.3.4 可知, $B(0,1)$+cone $\{(0,\ 1)^{\mathrm{T}}\}$ 为系统 (9.3.9) 的生存域. 显然, $B(0,1)$+cone $\{(0,\ 1)^{\mathrm{T}}\}$ 既非椭球也非多面体形状的.

参考文献

[1] Aubin J P. Further properties of Lagrange multipliers in nonsmooth optimization[J]. Applied Mathematics and Optimization, 1980, 6:79~90.

[2] Aubin J P, Ekeland I. Applied Nonlinear Analysis[M]. New York: John Wiley & Sons, 1984.

[3] Aubin J P, Frankowska H. Set-valued Analysis[M]. Boston: Birkhauser, 1990.

[4] Aubin J P. Viability Theory[M]. Boston:Birkhauser, 1991.

[5] Aubin J P, Lygeros J, Quincampoix M, Shankar S, et al.. Impulse differential inclusions: a viability approach to hybrid systems[J]. IEEE Trans Automatic Control, 2002, 47(1):2~20.

[6] Blanchini F. Set invariance in control[J]. Automatica, 1999, 35(11):1747~1767.

[7] Bonnans, J F, Shapiro A. Perturbation Analysis of Optimization Problems[M]. New York: Springer-Verlag, 2000.

[8] Caprari E, Penot J P. Tangentially ds functions[J]. Optimization, 2007, 56(1, 2):25~38.

[9] Chen GY, Goh C J, Yang X Q. The gap function of a convex multicriteria optimization problem[J]. European Journal of Operational Research, 1998, 111(1):142~151.

[10] Chaney R W. Piecewise C^k function in nonsmooth analysis[J]. Nonlinear Analysis, 1990, 15:649~660.

[11] Clarke F H. Generalized gradients and applications[J].Trans. Amer. Math Soc., 1975, 205:247~262.

[12] Clarke F H, Leda Yu S, Stern R J, et al.. Nonsmooth Analysis and Control Theory[M]. New York: Springer-Verlag, 1998.

[13] Clarke F H. Optimization and Nonsmooth Analysis[M]. New York: John Wiley & Sons, 1983.

[14] Demyanov V F, Gamidow S, Sivelina T L. An algorithm for minimizing a certain class of quasidifferentiable functions[J]. Mathematical Programming Study, 1986, 29:74~84.

[15] Demyanov V F, Polyakova L N, Rubison A M. Nonsmooth and quasidifferentiability. Mathematical Programming Study[J]. 1986, 29:1~19.

[16] Demyanov V F, Rubinov A M. Constructive Nonsmooth Analysis[M]. Frankfurt am Main: Peterlang, 1995.

[17] Demyanov V F, Stavroulakis G E, Polyakova L N, et al.. Quasidifferentiability and Nonsmooth Modelling in Mechanics, Engineering and Economic[M]. Dordercht: Kluwer Academic Publishers, 1996.

[18] Nguyen D, Tuan L A. Directional Kuhn-Tucker condition and duality for quasidifferentiable programs[J]. Acta Mathematica Vietnamica, 2003, 28:17~38.

[19] Dutta J. Generalized derivatives and nonsmooth optimization, a finite dimensional tour[J]. Sociedad de Estad'isticae Investigaci'on Operativa Top, 2005, 13(2):185~314.

[20] Eppler K, Luderer B. The Lagrange principle and quasidifferential calculus[J]. TU Karl-Marx-Stadt, 1987, 29:187~192.

[21] Feng E M, Wang Y. On the use of ABS algorithms in the modeling of oil deposits[J]. Ricerca Operative, 2002, 31(8):99~100.

[22] Fukushima M, Qi L. A global and superlinearly convergent algorithm for nonsmooth convex optimization[J]. SIAM J Optimization, 1996, 6:1106~1120.

[23] Gao D Y. Dual extremum principles in finite deformation theory with applications in post-buckling analysis of nonlinear beam model[J]. Applied Mechanics Review. ASME, 1997, 50:64~71.

[24] Gao D Y. Duality Principles in Nonconvex Systems: Theory, Methods and Applications[M]. Boston: Kluwer Academic Publishers, 2000.

[25] Gao Y. Calculating the proximal subdifferential via the quasidifferential[J]. Applied Mathematics Letters. to appear.

[26] Gao Y, Lygeros J, Quincampoix M. On the reachability problem of uncertain hybrid systems[J]. IEEE Trans Automatic Control, 2007, 52:1572~1586.

[27] Gao Y. Newton method of solving Karush-Kuhn-Tucker systems for a constrained minimax problem[J]. J Information and Optimization Sciences, 2007, 28:173~184.

[28] 高岩. 线性控制系统的生存域 [J]. 控制与决策, 2007, 22:833~835.

[29] Gao Y. Differences of polyhedra in matrix space and their applications to nonsmooth analysis[J]. J Optimization Theory and Applications, 2006, 130:429~440.

[30] Gao Y. Newton methods for quasidifferentiable equations and their convergence[J]. J Optimization Theory and Applications, 2006, 131:417~428.

[31] Gao Y, Lygeros J, Quincampoix M. The reachability problem for uncertain hybrid systems revisited: a viability theory perspective[C]. Lecture Notes in Computer Science. Springer-Verlag, 2006, 3927:242~256.

[32] 高岩. 一类非线性控制系统关于非光滑区域生存性的判别 [J]. 控制与决策, 2006, 21(8): 923~925.

[33] 高岩. 一类非光滑优化及其在控制系统稳定化中的应用 [J]. 控制与决策, 2006, 21(1): 118~120.

[34] Gao Y. Representation of the Clarke generalized Jacobian via the quasidifferential[J]. J Optimization Theory and Applications, 2004, 123:519~532.

[35] Gao Y, Li X W. Nonsmooth equations approach to a constrained minimax problem[J]. Applications of Mathematics, 2005, 50(2):115~130.

[36] Gao Y. Representative of quasidifferentials and its formula for a quasidifferentiable function[J]. Set-Valued Analysis, 2005, 13(4):323～336.

[37] Gao Y, Lygeros J, Quincampoix M, Seube N. On the control of uncertain impulsive system: approximate stabilisation and controlled invariance[J]. International J Control, 2004, 77(16):1393～1407.

[38] Gao Y. Representation of the Clarke generalized Jacobian via the quasidifferential[J]. J Optimization Theory and Applications, 2004, 123(3):519～532.

[39] Gao Y, Lygeros J, Quincampoix M, et al.. Approximate stabilisation of uncertain hybrid systems[C]. Lecture Notes in Computer Science, 2003, 2623:203～215.

[40] Gao Y. Newton methods for solving nonsmooth equations via a new subdifferential[J]. Mathematical Methods of Operations Research, 2001, 54:239～257.

[41] Gao Y. Calculating an element of B-differential for a vector-valued maximum function[J]. Numerical Functional Analysis and Optimization, 2001, 22(5, 6):561～572.

[42] Gao Y, Xia Z Q, Zhang L W. Kernelled quasidifferential for a quasidifferentiable function in two-dimensional space[J]. J Convex Analysis, 2001, 8(2):401～408.

[43] Gao Y. Newton methods for solving two classes of nonsmooth equations[J]. Applications of Mathematics, 2001, 46(3):215～229.

[44] Gao Y. Demyanov difference of two sets and optimality conditions of Lagrange multiplier type for constrained quasidifferentiable optimization[J]. J Optimization Theory and Applications, 2000, 104(2):377～394.

[45] Gao Y. A modified algorithm of finding an element of Clarke generalized gradient for a smooth composition of max-type functions[J]. J Computational Mathematics, 2000, 18(5):513～520.

[46] 高岩. 极大值复合函数 Clarke 广义梯度计算的一个新方法 [J]. 应用数学学报, 2000, 23(1):15～20.

[47] Grzybowski J. Minimal quasidifferential of a piecewise linear function[J]. Zeitschrift fur Analysis und Ihreanwendungen, 2005, 24(1):189～202.

[48] Han J Y, Liu G S, Wang S Y. A new descent algorithm for solving quadratic bilevel programming problem[J]. Acta Mathematicae Applicatae Sinica (English Series), 2000, 16(3):235～244.

[49] Harker P T, Pang J S. Finite-dimensional variational inequality and nonlinear complementarity problems: a survey of theorem, algorithm and applications[J]. Mathematical Programming, 1990, 56:161～220.

[50] Hiriart-Urruty J B, Lemarechal C. Convex Analysis and Minimization Algorithms[M]. Berlin: Springer-Verlag, 1993.

[51] Huang Z H. Generalization of an existence theorem for variational inequalities[J]. J Optimization Theory and Applications, 2003, 118(3):567～585.

[52] 胡毓达, 孟志青. 凸分析与非光滑分析 [M]. 上海: 上海科学技术出版社, 2000.

[53] Ioffe A D. Nonsmooth analysis: differential calculus of nondifferentiable mappings[J]. Tran of the American Mathematical Society, 1981, 266:1~56.

[54] Ioffe A D. Necessary and sufficient conditions for local minimum. 1: a reduction theorem and first order conditions[J]. SIAM J Control and Optimization, 1979, 17(2): 245~250.

[55] Kelley J E. The cutting plane method for solving convex programs[J]. J SIAM, 1960, 8:703~712.

[56] Kiwiel K C. Methods of Descent for Nondifferentiable Optimization: Lecture Notes in Mathematics 1133[M]. Berlin: Springer-Verlag, 1985.

[57] Lemaréchal C, Sagastizabal C. Variable metric bundle methods: from conceptual to implementable forms[J]. Mathematical Programming, 1997, 76(3):393~410.

[58] Li D H, Fukushima M, Qi L, et al.. Regularized Newton methods for convex minimization problems with singular solutions[J]. Computational Optimization and Applications, 2004, 28:131~147.

[59] Li Z F, Wang S Y. ε-approximate solutions in multiobjective optimization[J]. Optimization, 1998, 44(2):161~174.

[60] 刘光中. 凸分析与极值问题 [M]. 北京: 高等教育出版社, 1991.

[61] Liu J, Gao Y. A subalgorithm for quasidifferentiable equations[J]. International J Pure and Applied Mathematics, 2005, 23(3):335~342.

[62] Li, D, Sun X L, Biswal M P, et al.. Convexification, concavification, and monotonization in global optimization[J]. Annals of Operations Research, 2001, 105:213~226.

[63] Luderer B. directional derivative estimates for the optimal value function of a quasidifferentiable programming problem[J]. Mathematical Programming, 1991, 51:333~348.

[64] Luderer B, Weigelt J. A solution method for a special class of nondifferentiable unconstrained optimization problems[J]. Computational Optimization and Applications, 2003, 24:83~93.

[65] Luksan L, Vlcek J. A bundle Newton method for nonsmooth unconstrained minimization[J]. Mathematical Programming, 1998, 83:373~391.

[66] Milani B E A, Dorea C E T. On invariant polyhedra of continuous-time linear systems subject to additive disturbances[J]. Automatica, 1996, 32(5):785~789.

[67] Mifflin R. Semismooth and semiconvex functions in constrained optimization. SIAM J Control and Optimization, 1977, 15:959~972.

[68] 欧宜贵, 廖琳武. 不可微规划的理论简介 [J]. 海南大学学报 (自然科学版), 1999, 17: 380~382.

[69] Pang J S, Ralph D. Piecewise smoothness, local invertibility and parametric analysis of normal maps[J]. Mathematics of Operations Research, 1996, 21:401~426.

[70] Pallaschke D and Urbanski R.. Pairs of Compact Convex Sets-Fractional Arithmetic with Convex Set[M]. New York: Kluwer Publisher, 2002.

[71] Polyakova L N. On the minimization of a quasidifferentiable function subject to equality-type constraints[J]. Mathematical Programming Study, 1986, 29:44～55.

[72] Pickl S. Solving the semismooth equivalence problem[J]. European J Operational Research, 2004, 157:8～78.

[73] Pu D, Tian W. Globally convergent inexact generalized Newton's methods for nonsmooth equations[J]. J Computing and Applied Mathematics, 2002, 138:37～49.

[74] Qi L. Convergence analysis of some algorithm for solving nonsmooth equations[J]. Mathematics of Operations Research, 1993, 18:227～244.

[75] Qi L, Sun J. A nonsmooth version of Newton's method[J]. Mathematical Programming, 1993, 58:353～367.

[76] Quincampoix M, Seube N. Stabilization of uncertain control systems through piecewise constant feedback[J]. J Mathematics Analysis and Applications, 1998, 218(1): 240～255.

[77] Rifford L. Existence of Lipschitz and semiconcave control-Lyapunov function[J]. SIAM J Control and Optimization, 2000, 39:1043～1064.

[78] Robinson S M. Generalized equations and their solutions, Part I: basic theory[J]. Mathematical Programming Study, 1979, 10:128～141.

[79] Rockafellar R T. Convex Analysis[M]. New Jersey: Princeton University Press, 1970.

[80] Rockafellar R T, Wets R J B. Variational Analysis[M]. New York: Springer-Verlag, 1998.

[81] Rubinov A M, Akhundov I S. *Difference of convex compact sets in the sense of Demyanov* and its applications in nonsmooth analysis[J]. Optimization, 1992, 23:179～188.

[82] Shapiro A. On optimality conditions in quasidifferentiable optimization[J]. SIAM J Control and Optimization, 1984, 22:610～617.

[83] 史树中. 凸分析 [M]. 上海: 上海科学技术出版社, 1990.

[84] Shor N Z. *Minimization methods for nondifferentiable functions*[J]. *Berlin: Springer-*Verlag, 1985.

[85] Šmietańki M J. An approximate Newton method for non-smooth equations with finite max functions[J]. Numerical Algorithms, 2006, 41(3):219～238.

[86] Šmietańki M J. A generalized Jacobian based Newton method for semismooth block-triangular system of equations[J]. J Computational and Applied Mathematics, 2007, 205(1):305～313.

[87] Solodov M V. On approximations with finite precision in bundle methods for nonsmooth optimization[J]. J Optimization Theory and Applications, 2003, 119:151～165.

[88] Song W. The solution set of a differential inclusion on a Banach space[J]. Applicationes Mathematicae, 1995, 23:13～23.

[89] 宋春玲, 夏尊铨, 张立卫. A note on the upper semi-continuity of Demyanov sum of quasidifferential mappings[J]. 运筹学学报, 2007, 11(1):33～38.

[90] Sun D, Han J. Newton and quasi-Newton methods for a class of nonsmooth equations and related problems[J]. SIAM J Optimization, 1997, 7:463～480.

[91] Uderzo A. Quasi-multipliers rules for quasidifferentiable extremum problems[J]. Optimization, 2002, 51(6):761～795.

[92] Vinter R B, Zheng H. Some finance problems solved with nonsmooth optimization techniques[J]. J Optimization Theory and Applications, 2003, 119(1):1～18.

[93] Wang W, Pang L P, Xia Z Q. The uv-decomposition on a class of D.C. functions and optimality conditions[J]. Acta Mathematicae Applicatae Sinica, 2007, 23(1):29～38.

[94] Wang S Y, Yamamoto Y, Yu M. A minimax rule for portfolio selection in frictional markets [J]. Mathematical Methods of Operations Research, 2003, 57(1):41～155.

[95] 王长钰, 韩继业. 非光滑半无限规划极大熵方法的稳定性 [J]. 中国科学 (A 辑), 1999, 29:593～599.

[96] Xia Z Q. On quasidifferential kernels [J]. Demonstratio Mathematica, 1993, XXVI: 159～182.

[97] Xia Z Q Finding Subgradient or descent direction of convex functions by external polyhedral approximation of subdifferentials[J]. Optimization Methods and Software, 1992, 1:273～295.

[98] Xia Z Q, Zhao C J. Second-order expansions for a class of quasidifferentiable functions[J]. Demonstratio Mathematica, 1989, 12:365～389.

[99] 修乃华, 韩继业. 对称锥互补问题 [J]. 数学进展, 2007, 36:1～12.

[100] Yang X Q, Jeyakumar V. Generalized second-order directional derivatives and optimization with $C^{1,1}$ functions[J]. Optimization, 1992, 26:165～185.

[101] Yang X M, Wang S Y. Near-subconvexlikeness in vector optimization with set valued functions[J]. Journal of Optimization Theory and Applications, 2001, 110:413～427.

[102] Yin H, Zhang K. The Fritz John type conditions in quasidifferentiable programming with equality constraints and their sufficiency[J]. Applied Mathematics, Journal of Chinese University, 1998, 13:99～104.

[103] Yin H X. An iterative method for general variation inequalities[J]. J Industrial and Management Optimization, 2005, 1(2):201～209.

[104] 殷洪友, 张可村. 具有等式约束的拟可微规划的 Fritz John 型条件及其充分性 [J]. 高校应用数学学报, 1998, 13:99～104.

[105] Yuan Y X. An example of only linearly convergence of trust region algorithms for nonsmooth optimization[J]. IMA J Numerical Analysis, 1984, 4:327～335.

[106] Yuan Y X. Conditions for convergence of trust region algorithms for nonsmooth optimization[J]. Mathematical Programming, 1985, 31:220～228.

[107] 袁亚湘, 孙文瑜. 最优化理论与方法 [M]. 北京: 科学出版社, 2003.

[108] Zhang L S. An algorithm for nondifferentiable optimization[J]. Acta Mathematicae Applicatae Sinica (English Series), 1988, 4(3):168～173.

[109] Zhang L W, Xia Z Q. Newton-type methods for quasidifferential equations[J]. Journal of Optimization Theory and Applications, 2001, 108:439~456.

[110] Zhang L W, Xia ZQ, Gao Y, Wang M Z. Star-kernel and star-differential in quasidifferential analysis[J]. Journal of Convex Analysis, 2002, 9(1):139~158.

[111] Zhao W H, Gao Y. Optimality conditions with quasidifferential for nonsmooth semidefinite programming[J]. World J Modelling and Simulation, 2006, 2(4):247~254.

[112] Zheng X Y, Yang X Q. Lagrange multipliers in nonsmooth semi-infinite optimization problems[J]. Mathematics of Operations Research, 2007, 32(1):168~181.

[113] Zhu D T. An affine scaling interior trust-region method for LC^1 minimization subject to bounds on variables[J]. Applied Mathematics and Computation, 2006, 172: 1272~1302.

《运筹与管理科学丛书》已出版书目

1. 非线性优化计算方法　袁亚湘　著　2008年2月
2. 博弈论与非线性分析　俞　建　著　2008年2月
3. 蚁群优化算法　马良等　著　2008年2月
4. 组合预测方法有效性理论及其应用　陈华友　著　2008年2月
5. 非光滑优化　高岩　著　2008年5月

[General Information]
书名=非光滑优化
作者=高岩著
页数=200
SS号=11987832
DX号=
出版日期=2008年05月第1版
出版社=科学出版社